





```

Query Match          75.8%      Score 73:  DB 10: Length 636;
Best Local Similarity 52.9%      Pred. No. 1.29e+01;
Matches          9;  Conservative          7;  Mismatches          1;  Indels          0;  Gaps          0;

Dc      1131  RMSNTTYTAAAYGARG 1147
      :::::|:|:|:|:|:|
Cp      17  ATCGCTCATTAATGAG 1
      S F I N E

RESULT      10
ID      045377      PRELIMINARY;      PRT;      2374 AA.
AC      045377;
DT      01-JUN-1998 (TREMBL,REL. 06, CREATED)
DT      01-NOV-1998 (TREMBL,REL. 08, LAST SEQUENCE UPDATE)
DT      01-JAN-1999 (TREMBL,REL. 09, LAST ANNOTATION UPDATE)
DE      F22G12.5 PROTEIN.
GN
OS      CAENORHABDITIS ELEGANS.
OC      EUAROTIA; METAEOA; NEMATODA; SECERNENTEAE; RHABDITIDA; RHABDITIDA;
NC      RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
[1]
RP      SEQUENCE FROM N.A.
RA      LEONARD N.;
RL      SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN      (2)
RP      SEQUENCE FROM N.A.
RX      MEDLINE: 94150718.
RA      WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKKS M.,
RA      BOFFIELD J., BUTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,
RA      CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA      GRADNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA      JONES M., KESHAW J., KIRSTEN T., LAISTER N., LATREILLE P.,
RA      LIGHTING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA      PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,

```

CC RHABDITININ; RHABDITIOIDEA; RHABDITIIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE; 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKES M.,  
RA BOFIELD J., BUSTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU 2., DUBREIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HARKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES W., KESSEAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFEEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMALDON N., SMITH A., SONNHAMMER E., SPADEN R., SUSTON J.,  
RA THERBY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPRAT J., WOLDMAN P.,  
RT #2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.;  
RL NATURE 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA MILLER N.;  
RL SUBMITTED (APR-1994) TO EMBL/GENBANK/DBDJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL SUBMITTED (APR-1994) TO EMBL/GENBANK/DBDJ DATA BANKS.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBDJ DATA BANKS.  
DR EMBL; U00046; G470361; -.  
DR PROSITE; PS01186; EGF\_2; 7.  
DR PfAM; PF00008; EGF; 4.

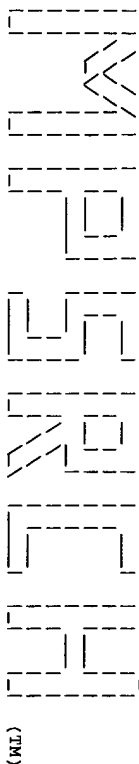












(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 12:39:30 1999; MasPar time 109.30 Seconds

Tabular output not generated. 407.328 Million cell updates/sec

Title: >US-09-103-287-5  
Description: (1-19) from US09103287.seq  
Perfect Score: 19  
N.A. Sequence: 1 CTTCAATATGACGATGC 19  
Comp: GAAGTAATTAATGCTGCTACG

Scoring table: TABLE default  
Gap 10

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 segs, 1171580779 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb1-est58  
1:em\_est10 2:em\_est11 3:em\_est17 4:em\_est18 5:em\_est2  
6:em\_est19 7:em\_gss1  
Database: genbank-est11  
8:gb\_est1 9:gb\_est10 10:gb\_est11 11:gb\_est12 12:gb\_est13  
13:gb\_est14 14:gb\_est15 15:gb\_est16 16:gb\_est17  
17:gb\_est18 18:gb\_est19 19:gb\_est20 20:gb\_est21  
21:gb\_est22 22:gb\_est23 23:gb\_est24 24:gb\_est25  
25:gb\_est26 26:gb\_est27 27:gb\_est28 28:gb\_est29  
29:gb\_est30 30:gb\_est31 31:gb\_est32 32:gb\_est33 33:gb\_est34  
34:gb\_est35 35:gb\_est36 36:gb\_est37 37:gb\_gss1 38:gb\_gss2  
39:gb\_gss3 40:gb\_gss4 41:gb\_gss5 42:gb\_gss6

Statistics: Mean 6.513; Variance 1.305; scale 4.992

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	Length	DB ID	Description	Pred. No.
1	84.2	698 39	A0029367	RPC111-37N19_TV RPC111	3.55e-03
2	78.9	359 26	A13844212	fb13a10.xl zebrafish c	7.13e-02
3	78.9	407 40	A0263297	CITB1-B1-2509G23.TR CI	7.13e-02
4	78.9	663 40	A0241183	RPC111-68N14_TV RPC111	7.13e-02
5	73.7	165 38	B95213	CIT-HSP-2172J8_TV CIT	1.25e+00
6	73.7	168 13	AA406838	MBAFC2560213 Brugia ma	1.25e+00
7	73.7	194 16	AA620120	v177c04.r1 Knowles Sol	1.25e+00
8	73.7	233 25	A1321054	d3c08nm.r1 Neurospora	1.25e+00
9	73.7	256 11	AA343898	EST44800 gall bladder	1.25e+00
10	73.7	318 8	Z44153	HSC1TF031 normalized 1	1.25e+00

c	11	14	73.7	319 24	A1267326	ag6f10.xl Stanley Fro	1.25e+00
c	12	14	73.7	355 12	AA389812	OS112 NaCl-treated Ara	1.25e+00
c	13	14	73.7	360 32	D73955	CELK067A9F Yuj1 Kohara	1.25e+00
c	14	14	73.7	367 40	A0231915	HS_3243.B1 A11 MR CIT	1.25e+00
c	15	14	73.7	380 37	FR0006589	F_rubripes GSS sequenc	1.25e+00
c	16	14	73.7	397 23	A1164965	A071P560 Hybrid aspen	1.25e+00
c	17	14	73.7	406 31	R65414	13918 Lambda-PRL2 Arab	1.25e+00
c	18	14	73.7	406 39	AQ132767	HS_3064.B2_H01_MF CIT	1.25e+00
c	19	14	73.7	412 39	AQ122186	HS_3083.A1_H05_MR CIT	1.25e+00
c	20	14	73.7	431 40	AQ141723	HS_3169.A2.B12_MR CIT	1.25e+00
c	21	14	73.7	435 40	AQ236429	HS_2041.A1_E12_MR CIT	1.25e+00
c	22	14	73.7	442 32	N44892	Yy12C06.r1 Soares mea	1.25e+00
c	23	14	73.7	459 41	AQ37265	HS_5009.A2.H12_SPEE RP	1.25e+00
c	24	14	73.7	463 19	R17201	Yg11f02.r1 Soares inf	1.25e+00
c	25	14	73.7	478 40	AQ242390	HS_2008.B1.A10_MR CIT	1.25e+00
c	26	14	73.7	480 19	T79899	Yd09e05.r1 Soares inf	1.25e+00
c	27	14	73.7	483 30	R60146	Yh12b12.r1 Soares inf	1.25e+00
c	28	14	73.7	487 39	AQ123302	HS_3090.B1.G08_TV CIT	1.25e+00
c	29	14	73.7	499 42	AQ384669	RPC111-141G3_TV RPC111	1.25e+00
c	30	14	73.7	499 33	W01347	za34f10.r1 Soares feta	1.25e+00
c	31	14	73.7	503 41	AQ295435	HS_3023.A2.E04_TV CIT	1.25e+00
c	32	14	73.7	508 40	AQ154678	HS_3034.B2.B01_TV CIT	1.25e+00
c	33	14	73.7	518 41	AQ370969	HS_3021.A1.D02_TV CIT	1.25e+00
c	34	14	73.7	520 37	FR0025366	F_rubripes GSS sequenc	1.25e+00
c	35	14	73.7	522 14	AA55363	CPEST.591 un12APC10WA	1.25e+00
c	36	14	73.7	531 27	A1464854	m292a02.Y1 Soares mous	1.25e+00
c	37	14	73.7	540 37	FR0006002	F_rubripes GSS sequenc	1.25e+00
c	38	14	73.7	550 36	AA138575	m08e10.r1 Soares mous	1.25e+00
c	39	14	73.7	585 16	C78161	C78161 Mouse 3.5-dpc b	1.25e+00
c	40	14	73.7	592 26	A0000400	n0000400 Bombyx mori p	1.25e+00
c	41	14	73.7	613 40	AQ255584	nbx00011F08r CUGI Rice	1.25e+00
c	42	14	73.7	663 40	AQ241183	RPC111-68N14_TV RPC111	1.25e+00
c	43	14	73.7	673 37	B71964	RPC111-7B22_TV RPC111	1.25e+00
c	44	14	73.7	674 42	AQ049790	nbx00001aet12f CUGI Ric	1.25e+00
c	45	14	73.7	737 41	AQ291390	nbx00038P14r CUGI Rice	1.25e+00

## ALIGNMENTS

RESULT	1	A0029367	698 bp	DNA	21-AUG-1998
LOCUS		RPC111-37N19_TV RPC111		Homo sapiens genomic clone R-37N19, genomic	
DEFINITION		survey sequence.			
ACCESSION		A0029367			
NTD		93274498			
VERSION		A0029367.1	GI:3274498		
KEYWORDS		GSS.			
SOURCE		human.			
ORGANISM		Homo sapiens			
		Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;			
		Primates; Catarrhini; Hominoidea; Homo.			
REFERENCE		1 (bases 1 to 698)			
AUTHORS		Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,			
		Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and			
		Venter,J.C.			
TITLE		Use of BAC End Sequences for Sequence-Ready Map Building (1998)			
JOURNAL		Unpublished (1998)			
COMMENT		Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: madams@tigr.org For clone availability, please contact Pieter de Jong (pieter@tigr.org, med.buifalo.edu). Clones may be purchased from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tigr/humgen/bac_end_search/bac_end_search.html. Seq primer: 77 Class: BAC ends. Location/Qualifiers 1. .698			

FEATURES  
Source



```

/organism="Homo sapiens"
/Note-Vector: PACes.6; Site_1: EcoRI; Site_2: EcoRI;
Rpci11 Human Male BAC Library"
/db_xref="taxon:9606"
/clone="R-37N19"
/clone.lib="Rpci11"
/sex="Male"
/cell_type="Lymphocytes"
BASE COUNT      246 a      112 c      113 g      227 t
ORIGIN
Query Match      84.2% Score 16; DB 39; Length 698;
Best Local Similarity 94.4%; Pred. No. 3.55e-03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db      260 TTCATTACTGACGATGC 277
      ||||| ||||| |||||
Oy      2 TTCATTATGAACGATGC 19

RESULT      2
LOCUS      A1384212      359 bp      mRNA      EST      27-JAN-1999
DEFINITION      fb13a10.x1 zebrafish fin day0 regeneration Danio rerio cDNA 3',
      mRNA sequence.
ACCESSION      A1384212
NID      94196993
VERSION      A1384212.1 GI:4196993
KEYWORDS      EST.
SOURCE      zebrafish.
ORGANISM      Danio rerio
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
      Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
      Cyprinidae; Cyprinidae; Rasbora; Danio.
REFERENCE      1 (bases 1 to 359)
      Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Maria,M.,
      Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
      Underwood,K., Steptoe,M., Theising,R., Allen,M., Bowers,Y.,
      Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
      Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
      Waterston,R. and Wilson,R.
      Mashu zebrafish EST Project 1998
      Unpublished (1998)
      On Jan 17, 1998 this sequence version replaced gi:1901040.

TITLE
JOURNAL
COMMENT

Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zdrafish@watson.wustl.edu
cDNA Library Preparation: Raymond Lee, cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
Resourcenzentrumprimatendank, Berlin, Germany (web address:
www.rzp.d.de)
Seq primer: T7 ET from Amersham.

FEATURES
source
1..359
Location/Qualifiers
/organism="Danio rerio"
/Note-Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; 1st
strand cDNA primed with (GA)10ACTGATCGAG(T)18, followed
by second strand synthesis, and ligated to 5' adapter
(5'-aattcgccagcag-3', 3'-gccgtgttc-5'. cDNA was cloned
directionally (EcoRI/XhoI) into Stratagene zap express
lambda phage arms. Mass in vivo excision done to obtain
inserts in pBK-CMV phagemid."
/db_xref="taxon:7955"
/map="875B10: 20: 20p11.23-20q11.23"
/clone.lib="zebrafish fin day0 regeneration"
/sex="mixed male and female"

/organism="Homo sapiens"
/Note-Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
/db_xref="taxon:9606"
/clone="2509G23"
/clone.lib="CITBI-El"
/sex="male"
/cell_type="sperm"
BASE COUNT      127 a      87 c      74 g      119 t
ORIGIN
Query Match      78.9% Score 15; DB 40; Length 407;
Best Local Similarity 94.1%; Pred. No. 7.13e-02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db      215 CTCATTATGAATGATGAT 231
      ||||| ||||| |||||
Oy      1 CTCATTATGAATGATGAT 17

RESULT      4
LOCUS      AQ241183      663 bp      DNA      GSS      29-SEP-1998
DEFINITION      Rpci11-68N14.TJ Rpci11 Homo sapiens genomic clone R-68N14, genomic
      survey sequence.
ACCESSION      AQ241183

/tissue_type="0 day fin regenerates"
/lab_host="E. coli XL0LR"
BASE COUNT      149 a      64 c      46 g      100 t
ORIGIN
Query Match      78.9% Score 15; DB 26; Length 359;
Best Local Similarity 89.5%; Pred. No. 7.13e-02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db      32 GCAACGTCATTATGAG 50
      ||| ||||| ||||| ||
Cp      19 GCATCGTCATTATGAG 1

RESULT      3
LOCUS      AQ263297      407 bp      DNA      GSS      26-OCT-1998
DEFINITION      CITBI-El-2509G23.FR CITBI-El Homo sapiens genomic clone 2509G23,
      genomic survey sequence.
ACCESSION      AQ263297.
NID      93791927
VERSION      AQ263297.1 GI:3791927
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
      Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
      Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 407)
      Adams,M.D., Rounsley,S.D., Zhao,S., Baas,S., Linher,K., Golden,K.,
      Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
      Venter,J.C.
      Use of a random human BAC End Sequence Database for Sequence-Ready
      Map Building
      Unpublished (1998)
      Other GSSs: CITBI-El-2509G23.TF
      Contact: Mark Adams
      Department of Eukaryotic Genomics
      The Institute for Genomic Research
      9712 Medical Center Dr., Rockville, MD 20850, USA
      Tel: 301 838 0200
      Fax: 301 838 0208
      Email: mdadams@tigr.org
      Clones are available from Research Genetics (info@resgen.com). BAC
      end search page:
      http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
      Seq primer: M13 Reverse
      Class: BAC ends.

FEATURES
source
1..407
Location/Qualifiers
/organism="Homo sapiens"
/Note-Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
/db_xref="taxon:9606"
/clone="2509G23"
/clone.lib="CITBI-El"
/sex="male"
/cell_type="sperm"
BASE COUNT      127 a      87 c      74 g      119 t
ORIGIN
Query Match      78.9% Score 15; DB 40; Length 407;
Best Local Similarity 94.1%; Pred. No. 7.13e-02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db      215 CTCATTATGAATGATGAT 231
      ||||| ||||| |||||
Oy      1 CTCATTATGAATGATGAT 17

RESULT      4
LOCUS      AQ241183      663 bp      DNA      GSS      29-SEP-1998
DEFINITION      Rpci11-68N14.TJ Rpci11 Homo sapiens genomic clone R-68N14, genomic
      survey sequence.
ACCESSION      AQ241183

```



NID 93668767  
 VERSION AQ241183.1 GI:3668767  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 663)  
 Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,  
 Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.,  
 Use of human BAC End Sequences for Sequence-Ready Map Building  
 Unpublished (1998)  
 TITLE Other\_GSSs: RPI11-68N14.TK  
 JOURNAL Contact: Mark Adams  
 COMMENT Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 For clone availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 Research Genetics (info@resgen.com). BAC end search page:  
 http://www.tigr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html.  
 Seq primer: Sp6  
 Class: BAC ends.

FEATURES  
 source  
 1..663  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 RPI11 Human Male BAC Library"  
 /db\_xref="taxon:9606"  
 /clone="R-68N14"  
 /clone\_lib="RPI11"  
 /sex="Male"  
 /cell\_type="lymphocytes"  
 BASE COUNT 281 a 125 c 103 g 151 t 1 others  
 ORIGIN

Query Match 78.9%; Score 15; DB 40; Length 663;  
 Best Local Similarity 94.1%; Pred. No. 7,13e-02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 561 ATCGTCATTATGAG 577  
 ||| ||||| ||||| |||||  
 Cp 17 ATCGTCATTATGAG 1

RESULT 5  
 LOCUS B95213 165 bp DNA GSS 25-JUN-1998  
 DEFINITION CIT-HSP-2172J8.TF CIT-HSP Homo sapiens genomic clone 2172J8,  
 genomic survey sequence.  
 ACCESSION B95213  
 NID 92977550  
 VERSION B95213.1 GI:2977550  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 165)  
 Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,  
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shiuuy,H.,  
 Simon,M. and Venter,J.C.  
 Use of a random BAC End Sequence Database for Sequence-Ready Map  
 Building  
 Unpublished (1997)  
 TITLE Contact: Mark Adams  
 JOURNAL Department of Eukaryotic Genomics  
 COMMENT The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200

Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:  
 http://www.tigr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html  
 Seq primer: M13-21;  
 Class: BAC ends.

FEATURES  
 source  
 1..165  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /note="Vector: pBeloBAC11; Site\_1: HindIII; Site\_2:  
 HindIII"  
 /db\_xref="taxon:9606"  
 /clone="2172J8"  
 /clone\_lib="CIT-HSP"  
 /sex="Male"  
 /cell\_type="Sperm"  
 BASE COUNT 53 a 28 c 35 g 49 t  
 ORIGIN

Query Match 73.7%; Score 14; DB 38; Length 165;  
 Best Local Similarity 93.8%; Pred. No. 1.25e+00;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 12 CAACGTCATTATGA 27  
 ||| ||||| ||||| |||||  
 Cp 18 CAACGTCATTATGA 3

RESULT 6  
 LOCUS AA406838 168 bp mRNA EST 01-MAY-1997  
 DEFINITION MBARC2502T3 Brugia malayi adult female CDNA (SAW95MLW-BMAF) Brugia  
 malayi CDNA clone AF2502 5', mRNA sequence.  
 ACCESSION AA406838  
 NID 92064817  
 VERSION AA406838.1 GI:2064817  
 KEYWORDS EST.  
 SOURCE Brugia malayi.  
 ORGANISM Brugia malayi  
 Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;  
 Filarioidea; Onchocercidae; Brugia.  
 1 (bases 1 to 168)  
 Blaxter,M.L., Waterfall,M., Daub,J., Lizotte,M., Baron,L. and  
 Jones,S.J.  
 Genes expressed in adult female Brugia malayi  
 Unpublished (1996)  
 COMMENT On Apr 14, 1993 this sequence version replaced gi:837511.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

CONTACT: Blaxter M.L.  
 Institute of Cell, Animal and Population Biology  
 University of Edinburgh  
 Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
 3UT, UK.  
 Tel: +44 131 650 6760  
 Fax: +44 131 670 5450  
 Email: mark.blaxter@ed.ac.uk  
 The ABI trace of this sequence can be viewed at  
 http://www.sanger.ac.uk/Brugia/AF2502T3.html This is the  
 full sequence of the CDNA clone. The polyA tail has been clipped  
 and is excluded from this sequence  
 Seq primer: T3.  
 Location/Qualifiers  
 1..168  
 /organism="Brugia malayi"  
 /note="Vector: Lambda Uni-ZAP XR; Site\_1: Eco RI; Site\_2:  
 Xho I; Lymphatic filarial nematode parasite of humans.  
 mRNA was prepared from approximately 50 adult females  
 isolated from the peritoneal cavity of jirds and  
 converted to double-stranded cDNA using reverse  
 transcriptase and oligo(dT) followed by RNase H and DNA  
 pol I. The library has 5 x 10E6 independent recombinants  
 and the average insert size is ~900bp. The library was  
 constructed by Michelle Lizotte-Waniewski. The



library is available from Dr.S.A.Williams, email:  
 genome@smith.edu.  
 /db\_xref="taxon:6279"  
 /clone="AFC25G02"  
 /clone\_lib="Brugia malayi adult female cDNA  
 (SAW96MLW-Bmarf)"  
 /sex="female"  
 /dev\_stage="adult"  
 /lab\_host="XLI-Blue MRF"  
 BASE COUNT 57 a 18 c 30 g 63 t  
 ORIGIN

Query Match 73.7%; Score 14; DB 13; Length 168;  
 Best Local Similarity 88.9%; Pred. No. 1.25e+00;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 39 CTCATTATGATGATGATG 56  
 |||||  
 1 CTCATTATGATGATGATG 18

RESULT 7  
 LOCUS AA620120 194 bp mRNA EST 09-OCT-1997  
 DEFINITION v177c04.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA  
 clone IMAGE:978246 5' similar to gb:M64241 QM PROTEIN (HUMAN);  
 gb:M93980 Mouse 24.6 kda protein mRNA, complete cds (MOUSE);, mRNA  
 sequence.  
 ACCESSION AA620120  
 NID 92523996  
 VERSION AA620120.1 GI:2523996  
 KEYWORDS EST  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 194)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Scheinberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 TITLE The Mashu-HMM Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1394964.

CONTACT: Marra M/Mouse EST Project  
 WashU-HMM Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LBNL; contact the  
 IMAGE Consortium (info@image.lbnl.gov) for further information.  
 MGI:558974  
 Trace considered overall poor quality  
 High quality sequence stop: 1.  
 Location/Qualifiers  
 1..194  
 /organism="Mus musculus"  
 /strain="B6D2 F1/J"  
 /note="Organ: embryo; Vector: pSPORT; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally from mRNA prepared  
 from 800 blastocysts. Primer: SalI(dT):  
 5'-CGGTGACCGTCGACCGTTTGT-3', CDNA were  
 cloned into the NotI/SalI sites of a pSPORT vector (Life  
 Technologies). Two different size selections: B1 (larger  
 inserts) and B3.  
 /db\_xref="taxon:10090"  
 /map="19q13.30"  
 /clone="IMAGE:978246"  
 /clone\_lib="Knowles Solter mouse blastocyst B1"  
 /tissue\_type="blastocyst"

/dev\_stage="embryo (pre-implantation)"  
 /lab\_host="DH10B"  
 BASE COUNT 52 a 47 c 43 g 52 t  
 ORIGIN

Query Match 73.7%; Score 14; DB 16; Length 194;  
 Best Local Similarity 93.8%; Pred. No. 1.25e+00;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 162 ATCGTTCAATATGAA 177  
 |||||  
 Cp 17 ATCGTTCAATATGAA 2

RESULT 8  
 LOCUS A1321054 233 bp mRNA EST 18-DEC-1998  
 DEFINITION d3c08nm.r1 Neurospora crassa morning cDNA library Neurospora crassa  
 cDNA clone d3c08nm 3', mRNA sequence.  
 ACCESSION A1321054  
 NID 94037036  
 VERSION A1321054.1 GI:4037036  
 KEYWORDS EST  
 SOURCE Neurospora crassa.  
 ORGANISM Neurospora crassa  
 Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;  
 Sordariales; Sordariaceae; Neurospora.  
 1 (bases 1 to 233)  
 Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.  
 TITLE Two Neurospora crassa EST Databases  
 JOURNAL Unpublished (1998)  
 COMMENT On Aug 21, 1998 this sequence version replaced.

CONTACT: Bruce A. Roe, University of Oklahoma, broe@ou.edu  
 Department of Chemistry and Biochemistry  
 Advanced Center for Genome Technology, University of Oklahoma  
 620 Parrington Oval, Norman, OK 73019, USA  
 Tel: 405 325 4912  
 Fax: 405 325 7762  
 Email: broe@ou.edu  
 We anticipate the future release of the cDNA clones to the Fungal  
 Genetics Stock Center  
 Seg primer: Universal Reverse Primer  
 High quality sequence stop: 118.  
 Location/Qualifiers  
 1..233  
 /organism="Neurospora crassa"  
 /strain="bd, frg7 A"  
 /note="Vector: pBluescript SK-, Site\_1: XbaI; Site\_2:  
 EcoRI; See: Bell-Pedersen,D., et al. PNAS 93:13096, 1996.  
 5' end of cDNA cloned into XbaI site of pBluescript;  
 3' end of cDNA cloned into EcoRI site of pBluescript"  
 /db\_xref="taxon:5141"  
 /map="X"  
 /clone="d3c08nm"  
 /clone\_lib="Neurospora crassa morning cDNA library"  
 /tissue\_type="tissue harvested following 22hr growth in  
 dark"  
 BASE COUNT 69 a 77 c 33 g 54 t  
 ORIGIN

Query Match 73.7%; Score 14; DB 25; Length 233;  
 Best Local Similarity 93.8%; Pred. No. 1.25e+00;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 8 TTCATTATGATGATGATG 23  
 |||||  
 Oy 2 TTCATTATGATGATGATG 17

RESULT 9  
 LOCUS AA343898 256 bp mRNA EST 21-APR-1997  
 DEFINITION EST49800 Gall bladder I Homo sapiens cDNA 5' end, mRNA sequence.  
 ACCESSION AA343898



NID 01996188  
 VERSION AA343898.1 GI:1996188  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 256)

AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., White,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., Miller,C., Sutton,G., Blake,J.A., Brandon,R.C., Man'hai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fitchman,J.L., Geoghegan,N.S., Giodex,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Gao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrle,A., Fischer,C., Hastings,G.A., He,W.M., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Ventler,J.C.

JOURNAL Initial assessment of human gene diversity and expression patterns MEDLINE based upon 83 million nucleotides of cDNA sequence 96026280

COMMENT On Apr 14, 1993 this sequence version replaced gi:693512. Other\_Ests: TNC169726 Contact: Kerlavage, AR Bioinformatics The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850 USA Tel: 3018699056 Fax: 3018699423 Email: arkerlavage@tigr.org

FEATURES

Source

1. .256  
 /organism="Homo sapiens"  
 /note="Organ: gall bladder; Vector: pbluescript SK-;  
 Site\_1: EcoRI; Site\_2: XhoI"  
 /db\_xref="ATCC (inhost):145467"  
 /db\_xref="taxon:9606"  
 /map="15"  
 /clone.lib="gall bladder 1"  
 /sex="female"  
 /dev\_stage="adult, 25 yrs"  
 /dev\_stage="38 c 53 g 81 t 1 others

BASE COUNT 83 a 38 c 53 g 81 t 1 others

ORIGIN

Query Match 73.7%; Score 14; DB 11; Length 256;  
 Best Local Similarity 88.9%; Pred. No. 1.25e+00;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DB 98 CTCTTTAATGATGATG 115  
 |||||  
 1 CTTCATTAATGACGATG 18

RESULT 10  
 LOCUS 244153 318 bp mRNA EST 14-NOV-1994  
 DEFINITION HSC1TF031 normalized infant brain cDNA Homo sapiens cDNA clone c-1tf03, mRNA sequence.  
 ACCESSION 244153  
 NID 9573375  
 VERSION 244153.1 GI:573275  
 KEYWORDS EST.

SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 318)

AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes,M.D., Duprat,S., Houigatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pletu,G., Pouillot,Y., Sebastiani,Kakatchis,C. and Tessier,A.

TITLE IMAGE: molecular integration of the analysis of the human genome and its expression

JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995) MEDLINE 95277534

COMMENT Contact: Genethon Genexpress-Genethon Genethon Centre de recherche sur le Genome Humain 1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE Tel: 33169472800 Fax: 33160778698 Email: genexpress@genethon.fr Single read. Genexpress\_library\_idt: C; Genexpress\_sequence\_idt: ylc-1tf03 Seq primer: (21)M13.universal.

FEATURES

Source

1. .318  
 /organism="Homo sapiens"  
 /note="Organ: brain; Vector: lafmid BA; Site\_1: HindIII; Site\_2: NotI; sex:female; dev\_stage=3 months old; isolate=muscular atrophy patient; tissue\_type=total brain; total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA vector. Clone library from B.Souares, Psychiatry Dept. Columbia University, USA. Normalization method: Bento Soares, P.N.A.S in press"  
 /db\_xref="taxon:9606"  
 /map="16"  
 /clone.lib="normalized infant brain cDNA"  
 /sex="female"  
 /tissue\_type="total brain"  
 /dev\_stage="3 months old"  
 /dev\_stage="46 c 62 g 111 t 4 others

BASE COUNT 95 a 46 c 62 g 111 t 4 others

ORIGIN

Query Match 73.7%; Score 14; DB 8; Length 318;  
 Best Local Similarity 88.9%; Pred. No. 1.25e+00;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DB 280 CTCTTTAATGATGATG 297  
 |||||  
 1 CTTCATTAATGACGATG 18

RESULT 11  
 LOCUS A1267326 319 bp mRNA EST 17-NOV-1998  
 DEFINITION ag63f10.x1 Stanley Frontal SN pool 2 Homo sapiens cDNA clone IMAGE:2035627, mRNA sequence.  
 ACCESSION A1267326  
 NID 93886493  
 VERSION A1267326.1 GI:3886493  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 319)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucab,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R. WashU-NCI human EST Project Unpublished (1997)



COMMENT On Jan 19, 1998 this sequence version replaced gi:2150694.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Gibco.

FEATURES  
source

Location/Qualifiers

1..319

/organism="Homo sapiens"

/note="Organ: Brain; Vector: PCR2.1-TOPO (Invitrogen);  
Site: 1: ECORI; Total RNA (purified with RNeasy and DNase  
before use) was reverse transcribed using a modified  
oligo-dT primer containing RsaI and HindIII sites.  
Double-stranded cDNA was digested with RsaI, resulting in  
blunt ended cDNA of an average 0.1-2 kb in length.  
Digested cDNA was split into two sets, one used as is as  
the driver, the other set was split in half again and each  
half linked to a different adaptor  
(5'-TCGACCGCGCGCGCGCGCGT-3' or 5'-  
ACGCGCTGTCGCGCGCGCGCGT-3'), to be used as tester.  
Subtraction was performed using the Clontech PCR Select  
cDNA subtraction kit. Pool of two schizophrenia, male age  
44 and female age 56 (S-116, S-118) subtracted by pool of  
two mentally normal male individuals ages 41 and 53  
(S-124, S-141). Tissues were obtained from the Stanley  
Neuropathology Consortium (www.stanleylab.org). Library  
constructed and subtracted by Dr. Nancy Johnston [(410)  
614-3918, nj@welchlink.welch.jhu.edu]."  
/db\_xref="taxon:9606"  
/map="21q"  
/clone="IMAGE:2035627"  
/clone\_lib="Stanley Frontal SN pool 2"  
/tissue\_type="Frontal lobe (see description)"  
/lab\_host="DHI08 (phage-resistant)"

BASE COUNT 98 a 53 c 67 g 100 t 1 others

ORIGIN

Query Match 73.7%; Score 14; DB 24; Length 319;  
Best Local Similarity 88.9%; Pred. No. 1.25e+00;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Db 204 CTTCTTAATGATGATG 221  
||||| ||||| ||||| |||||  
QY 1 CTTCTTAATGACGATG 18

RESULT 12  
LOCUS AA389812 355 bp mRNA EST 23-APR-1997  
DEFINITION OS172 NaCl-treated Arabidopsis subtraction library Arabidopsis  
thaliana cDNA 5', mRNA sequence.  
ACCESSION AA389812  
NID 92042798  
VERSION AA389812.1 GI:2042798  
KEYWORDS EST.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Rosidae; Capareales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 355)  
Pih.K.T., Park.J.M., Jang.H.J., Kang.S.G., Piao.H.L. and Hwang,I.  
EST of salt inducible mRNA in Arabidopsis thaliana  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1394805.

Contact: Kyeong Tae Pih  
Department of Plant Molecular Biology  
Gyeongsang National Univ., Plant Molecular Biology and

Biotechnology Research Center  
Ulsju, Kyungnam 660-701, Korea  
Tel: 82-591-751-5193  
Fax: 82-591-759-9363  
Email: ihwang@nongae.gsnu.ac.kr  
Seq primer: T3  
High quality sequence stop: 278.

FEATURES  
source

Location/Qualifiers

1..355

/organism="Arabidopsis thaliana"

/strain="Columbia"

/note="Subtracted cDNA library from salt(NaCl)-treated  
whole plants"

/db\_xref="taxon:3702"

/clone\_lib="NaCl-treated Arabidopsis subtraction library"

/dev\_stage="7 day-old"

BASE COUNT 98 a 80 c 80 g 93 t 4 others

ORIGIN

Query Match 73.7%; Score 14; DB 12; Length 355;  
Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 36 CTTCTTAATGAG 49  
||||| ||||| ||||| |||||  
Cp 14 GTTCTTAATGAG 1

RESULT 13  
LOCUS D73955 360 bp mRNA EST 14-DEC-1995  
DEFINITION CEK067A9F Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA  
clone yk67a9 5', mRNA sequence.  
ACCESSION D73955  
NID 91119740  
VERSION D73955.1 GI:1119740  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;  
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 360)  
Kohara.Y., Mitsuki.H., Nishigaki.A., Mochizuki.T., Sugimoto.A. and  
Tabara.H.  
Toward an expression map of the C.elegans genome  
Unpublished (1994)  
On Apr.14, 1993 this sequence version replaced gi:837965.

Contact: Yuj1 Kohara  
Gene Library Lab  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 0559-75-0771  
Fax: 0559-75-6240  
Email: yk67a9db.nig.ac.jp  
High quality sequence stop: 273.

FEATURES  
source

Location/Qualifiers

1..360

/organism="Caenorhabditis elegans"

/strain="CB1489 him-8(e1489)"

/note="dev\_stage=varied; sex=Hermaphrodite male;  
tissue\_type=whole animal"

/db\_xref="taxon:6239"

/clone\_lib="Yuj1 Kohara unpublished cDNA"

/clone\_lib="yk67a9"

/clone\_lib="Yuj1 Kohara unpublished cDNA"

BASE COUNT 110 a 71 c 77 g 101 t 1 others

ORIGIN

Query Match 73.7%; Score 14; DB 32; Length 360;  
Best Local Similarity 88.9%; Pred. No. 1.25e+00;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 52 CTTCTTAATGACGCTG 69  
||||| ||||| ||||| |||||



QY 1 CTCATTATGACGATG 18

RESULT 14  
LOCUS AQ231915 367 bp DNA GSS 25-SEP-1998  
DEFINITION HS\_3243\_B1\_A11\_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3243 Col-21 Row-B, genomic survey sequence.

ACCESSION AQ231915  
NID 93657144  
VERSION AQ231915.1 GI:3657144  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 367)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J., Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.  
Construction of a Characterized Clone Resource for Genomic Sequencing  
Unpublished (1998)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3243 row: B column: 21  
Class: BAC ends  
High quality sequence stop: 367.  
Location/Qualifiers  
1..367  
/organism="Homo sapiens"  
/note="Organ: Sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"  
/db\_xref="taxon:9606"  
/clone\_lib="Plate-3243 Col-21 Row-B"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"

BASE COUNT 116 a 46 c 90 g 115 t  
ORIGIN

Query Match 73.7%; Score 14; DB 40; Length 367;  
Best Local Similarity 100.0%; Pred. No. 1.25e+00; Mismatches 0; Indels 0; Gaps 0;  
Matches 14; Conservative

Db 286 CATTAATGACGAT 299  
|||||  
QY 4 CATTAATGACGAT 17

RESULT 15  
LOCUS FR0006589 380 bp DNA GSS 02-MAR-1997  
DEFINITION F.rubripes GSS sequence, clone 122014BD2, genomic survey sequence.  
ACCESSION 290399  
NID 91867613  
VERSION 290399.1 GI:1867613  
KEYWORDS GSS; genome survey sequence.  
SOURCE Fugu rubripes.  
ORGANISM Fugu rubripes.  
Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.  
1 (bases 1 to 380)  
Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umranta,Y., Williams,G. and Brenner,S.  
Direct Submission  
Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hmp.mrc.ac.uk

COMMENT Vector: pBluescript II KS  
V\_type: phagemid  
PRIMER: KS  
DESCR: One pass dye-terminator sequencing of cosmid cloned genomic sequence.  
FEATURES Location/Qualifiers  
source 1..380  
/organism="Fugu rubripes"  
/db\_xref="taxon:3103"  
/clone\_lib="cosmid 122014"  
/clone="122014BD2"

BASE COUNT 94 a 105 c 75 g 88 t 18 others  
ORIGIN

Query Match 73.7%; Score 14; DB 37; Length 380;  
Best Local Similarity 93.8%; Pred. No. 1.25e+00; Mismatches 1; Indels 0; Gaps 0;  
Matches 15; Conservative


Db 357 TTCATTATGACGAT 372  
|||||  
QY 2 TTCATTATGACGAT 17

Search completed: Sat Nov 27 12:41:25 1999  
Job time : 115 secs.



**THIS PAGE BLANK (USPTO)**



\*\*\*\*\*  
  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Msrch\_un n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 12:49:22 1999; MasPar time 2429.08 Seconds

Tabular output not generated. 1541.672 Million cell updates/sec

Title: >US-09-103-287-1  
 (1-1351) from US09103287.seq  
 Description: 1351  
 Perfect Score: 1 ATGAGTAGAGCTTTATATAT.....TTATATATTTATATATAG 1351  
 N.A. Sequence: TACTCATCTCCACAAATATA.....AATATATCAAAATATATATCTC  
 Comp: 1

Scoring table: TABLE jmetric

Match STD : Dbase 0; Query 0

Searched: 646147 seqs, 138595363 bases x 2

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: emb158  
 1:em\_ba1 2:em\_ba2 3:em\_fun 4:em\_hlg 5:em\_hum1 6:em\_hum2  
 7:em\_in 8:em\_com 9:em\_or 10:em\_ov 11:em\_pat 12:em\_ph  
 13:em\_pl 14:em\_ro 15:em\_sts 16:em\_vi

Database: genbank111  
 17:gb\_ba1 18:gb\_ba2 19:gb\_hlg1 20:gb\_hlg2 21:gb\_in1  
 22:gb\_in2 23:gb\_com 24:gb\_ov 25:gb\_pat 26:gb\_ph 27:gb\_pl1  
 28:gb\_pl2 29:gb\_pl3 30:gb\_pr2 31:gb\_pr3 32:gb\_ro  
 33:gb\_st 34:gb\_sts 35:gb\_sy 36:gb\_un 37:gb\_vi

Statistics: Mean 9.299; Variance 3.259; scale 2.853

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	867	64.2	1314	17	AF034076	Staphylococcus aureus
2	21	1.6	31281	21	CEC01F6	Caenorhabditis elegans
3	20	1.5	2322	17	ECU73760	Escherichia coli dehyd
4	20	1.5	4043	17	BACSKW01	Bacillus thuringiensis
5	20	1.5	5570	24	OMTIGGEN	Oreochromis mossambicu
6	20	1.5	5848	29	AB011134	Homo sapiens mRNA for
7	20	1.5	10036	18	AE000299	Escherichia coli K-12
8	20	1.5	11978	18	U67574	Methanococcus jannasch
9	20	1.5	12188	18	U32809	Haemophilus influenzae
10	20	1.5	17913	17	D90848	E.coli genomic DNA, Ko
11	20	1.5	47045	21	CEC17H12	Caenorhabditis elegans
12	20	1.5	209071	22	AC005285	Drosophila melanogaste
13	20	1.5	211037	19	AC005537	Homo sapiens clone GS3

14	19	1.4	524	34	G09531	human STS CHIC.GCT3B06	1.46e+01
15	19	1.4	2650	29	HUMICV3BA	Human germ-line IGH Cha	1.46e+01
16	19	1.4	3840	27	CHEGRPL	E.gracilis DNA for rib	1.46e+01
17	19	1.4	4435	27	SCRASR1	Yeast CN31C chromosome	1.46e+01
18	19	1.4	4597	21	PESC04106	Plasmodium falciparum	1.46e+01
19	19	1.4	9672	37	BSMRYP	Brome streak mosaic ry	1.46e+01
20	19	1.4	12361	18	U32787	Haemophilus influenzae	1.46e+01
21	19	1.4	14459	22	AE001383	Plasmodium falciparum	1.46e+01
22	19	1.4	34514	22	CEIC25B8	Caenorhabditis elegans	1.46e+01
23	19	1.4	36539	21	CEK08H2	Caenorhabditis elegans	1.46e+01
24	19	1.4	37322	30	HSU237H1	Human DNA sequence firo	1.46e+01
25	19	1.4	41017	27	CHEG2	Englema gracilis Z chl	1.46e+01
26	19	1.4	41275	21	CELP47D12	Caenorhabditis elegans	1.46e+01
27	19	1.4	74192	20	AC005504	Plasmodium falciparum	1.46e+01
28	19	1.4	76618	21	AC004657	Drosophila melanogaste	1.46e+01
29	19	1.4	80531	19	HSV46E01	Human DNA sequence ***	1.46e+01
30	19	1.4	103361	19	AC006429	Arabidopsis thaliana	1.46e+01
31	19	1.4	106455	19	AC004913	Homo sapiens clone DJO	1.46e+01
32	19	1.4	110000	19	PFMAL4P1	Plasmodium falciparum	1.46e+01
33	19	1.4	110000	19	CEY43D4	Caenorhabditis elegans	1.46e+01
34	19	1.4	129098	29	HS46H23	Human DNA sequence firo	1.46e+01
35	19	1.4	139316	27	ATF7H19	Arabidopsis thaliana D	1.46e+01
36	19	1.4	143172	27	CLEGGCA	E.gracilis chloroplast	1.46e+01
37	19	1.4	148020	19	HS796E4	Human DNA sequence ***	1.46e+01
38	19	1.4	151052	20	AC005506	Plasmodium falciparum	1.46e+01
39	19	1.4	186555	31	AC006313	Homo sapiens chromosome	1.46e+01
40	19	1.4	268531	19	CEY51H4	Caenorhabditis elegans	1.46e+01
41	19	1.4	279110	17	RPX03	Rickettsia prowseki	1.46e+01
42	19	1.4	315339	27	SCCHRTT	S.cervisiae chromosome	1.46e+01
43	18	1.3	3322	34	HUMC5532	Human chromosome 5 LAN	6.68e+01
44	18	1.3	2363	35	CVU4939	Cloning vector pAM120,	6.68e+01
45	18	1.3	118595	29	HS288L1	Human DNA sequence firo	6.68e+01

## ALIGNMENTS

RESULT LOCUS	1	AF034076	1314 bp	DNA	BCT	26-NOV-1997
DEFINITION		Staphylococcus aureus UDP-N-acetylmuuramoyl-L-alanine synthetase				
ACCESSION		AF034076				
NID		92642658				
VERSION		AF034076.1	GI:2642658			
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						
gene						
CDS						

Location/Qualifiers  
 1..1314  
 /organism="Staphylococcus aureus"  
 /db\_xref="taxon:1280"  
 1..1314  
 /gene="murC"  
 1..1314  
 /gene="murC"  
 /function="cell wall biosynthesis"  
 /note="MurC: UDP-N-acetylmuramate-alanine ligase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="UDP-N-acetylmuramoyl-L-alanine synthetase"  
 /protein\_id="AAB87090.1"



```

/db_xref="pid:92642659"
/translaton="MTHYFVIGKSGMSLQIMHDLHEVOGSDIENYVTEVALR
NGKIKLPDANNIKEDMVIQGNAPASSHEEKARAHOKMLDVSYNDLQIIDDYT
SVAVGAKSTSTGLSHVNGDKSTIGSLGSPESDYFPEACEYRHFSL
YKEDVAMTNIDHDHDFKIDNDFDAROMAHNKKGIANGDEDEHRIKEDVPI
YYGFSDSDDIYQNOIOTDKRPAFYVDYDEYDFLSPOYDHYLVNAIYAIYSY
LEKLDVNTIKREALTEYGVKRRNETTIAQVIVDVAHHPREISATIEARKKYPK
EVAVFQPTFSTQAFLENFASLSKARVFCETFGSIRENTGALTOTDLIDKIEG
ASLINDSDIVLEOFDPAVAVLFEKAGDIDKIQNAYVILKIKMKNAF"
BASE COUNT      467 a      163 c      250 g      434 t
ORIGIN
Query Match      64.2% Score 867; DB 17; Length 1314;
Best Local Similarity 99.3%; Pred. No. 0.00e+00;
Matches 1107; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Db      1 ATGCACATCATCTTTGTCGGAATTAAGTCTGCGCATGATTCATAGCACAATC 60
      22 ATGCACATCATCTTTGTCGGAATTAAGTCTGCGCATGATTCATAGCACAATC 81
      61 ATGCATATTAGACATGAGATCAAGATCGGATATTGAGAACTAGTATTACAGAA 120
      82 ATGATATATTAGACATGAGATCAAGATCGGATATTGAGAACTAGTATTACAGAA 141
      121 GTTGCTCTTGAATAAGGGATTAATAATTTACATTGATGCTAATAACATAAAGAA 180
      142 GTTGCTCTTGAATAAGGGATTAATAATTTACATTGATGCTAATAACATAAAGAA 201
      181 GATATGATGATTATACAGAGTAATGATCGGATCGGATCGGATCGGATCGGATCGG 240
      202 GATATGATGATTATACAGAGTAATGATCGGATCGGATCGGATCGGATCGGATCGG 261
      241 CATCAATGAATTAAGTCTTGAATTAATAATTTTATAGACAGATTTATGATCAA 300
      262 CATCAATGAATTAAGTCTTGAATTAATAATTTTATAGACAGATTTATGATCAA 321
      301 TATACCTAGTACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
      322 TATACCTAGTACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 381
      361 CATGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
      382 CATGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 441
      421 TTGCGTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
      442 TTGCGTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 501
      481 TATTAACCTGATTAAGCAATTAATGATGATGATGATGATGATGATGATGATG 540
      502 TATTAACCTGATTAAGCAATTAATGATGATGATGATGATGATGATGATGATG 561
      541 GATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
      562 GATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 621
      601 ATTGCTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
      622 ATTGCTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 681
      661 TATGATTTAAGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
      682 TATGATTTAAGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATG 741
      721 ACTGCTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
      742 ACTGCTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 801
      781 GGTGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
      802 GGTGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 861

```

```

Db      841 GATGTTACCAATTTAAGACGATTAAGACGTTGGTGGTGTAAAGCTGTTCAAT 900
      862 GATGTTACCAATTTAAGACGATTAAGACGTTGGTGGTGTAAAGCTGTTCAAT 921
      901 GAAACATCAATTTCAATCAAGTATTGATGATGATGATGATGATGATGATGATGATG 960
      922 GAAACATCAATTTCAATCAAGTATTGATGATGATGATGATGATGATGATGATGATG 981
      961 AGTGCTACATTTAAGACGATTAAGACGATTAAGACGTTGGTGGTGTAAAGCTGTT 1020
      982 AGTGCTACATTTAAGACGATTAAGACGATTAAGACGTTGGTGGTGTAAAGCTGTT 1041
      1021 CAACACACACCTTCTCTAGACACACGATTTTAAAGATTTGACAGAAATTAAGT 1080
      1042 CAACACACACCTTCTCTAGACACACGATTTTAAAGATTTGACAGAAATTAAGT 1101
      1081 AAAGCAGATGCTGATTTATGATGATGATGATGATGATGATGATGATGATGATG 1140
      1102 AAAGCAGATGCTGATTTATGATGATGATGATGATGATGATGATGATGATGATG 1161

```

```

RESULT      2
LOCUS      CEC01F6      31281 bp      DNA
DEFINITION      Caenorhabditis elegans cosmid C01F6, complete sequence.
ACCESSION      Z68213
VERSION      268213.1
KEYWORDS      HTG; Cytochrome P450; FEM-3; Sex-determining protein.
SOURCE      Caenorhabditis elegans.
ORGANISM      Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditi; Rhabditidae;
      Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
      1 (bases 1 to 31281)
REFERENCE
AUTHORS      Biers, M.
TITLE      Direct Submission
JOURNAL      Submitted (13-DEC-1995) Louis, MO 63110, USA. E-mail:
      jesssanger@ac.uk or twenematode.wustl.edu
      2 (bases 1 to 31281)

```

```

REFERENCE
AUTHORS      Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,
      Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
      Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
      Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
      Johnston, L., Jones, M., Kershav, J., Kirsten, J., Laister, N.,
      Latreille, P., Lightning, J., Lloyd, C., McMurtry, A., Mortimore, B.,
      O'Callaghan, M., Parsons, J., Percy, C., Rifen, L., Roopra, A.,
      Saunders, D., Shownkeen, R., Smaildon, N., Smith, A., Sonhammer, E.,
      Staden, R., Sustain, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
      Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
      Wilkinson-Sproat, J., and Wohlman, P.
      2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
      elegans
      JOURNAL      Nature 368 (6466), 32-38 (1994)
      MEDLINE      94150718
      COMMENT      Coding sequences below are predicted from computer analysis, using
      predictions from GeneFINDER (P. Green, V. Washington), and other
      available information.
      For a graphical representation of this sequence and its analysis
      see:

```

http://webc.sanger.ac.uk/cgi-bin/display?db=wormbase&class=Sequence&object=C01F6

Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlap between sections once, or longer because we arrange for a small overlap between neighbouring submissions.

IMPORTANT: This sequence is not the entire insert of clone C01F6. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.



The true left end of clone C01F6 is at 1 in this sequence. The true right end of clone C01F6 is at 9279 in sequence 282266. The true right end of clone ZC410 is at 2893 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence 268270. The end of this sequence (31178..31281) overlaps with the start of sequence 282266.

## FEATURES

source

1..31281

Location/Qualifiers

/organism="Caenorhabditis elegans"

/db\_xref="taxon:6239"

/chromosome="IV"

/complement(930..1254)

/gene="C01F6.9"

/complement(join(930..972,1021..1097,1153..1254))

/gene="C01F6.9"

/note="cDNA EST yk266c4.5 comes from this gene; cDNA EST yk266c4.3 comes from this gene"

/codon\_start=1

/protein\_id="CAA92441.1"

/db\_xref="PID:e1343696"

/db\_xref="PID:g3873821"

/db\_xref="GI:3873821"

/db\_xref="SPTREMBL:O62023"

/translation="MARCOOKIOSOONOKKADAAKAGIDGKAANAALNHCYCLAMPDPKTYKOHENKHPKSPALVELA"

1955..4099

/gene="C01F6.1"

join(1955..1992,2059..2244,2332..2818,2866..2969,3189..3435,3482..3609,3655..3817,3868..3945,3992..4099)

/gene="C01F6.1"

/note="cDNA EST EMBL:D35953 comes from this gene; cDNA EST EMBL:D33187 comes from this gene; cDNA EST yk445h11.3 comes from this gene; cDNA EST yk445h11.5 comes from this gene; cDNA EST yk353f12.3 comes from this gene; cDNA EST yk353f12.5 comes from this gene; cDNA EST yk310a7.3 comes from this gene; cDNA EST yk310a7.5 comes from this gene; cDNA EST yk280d11.3 comes from this gene; cDNA EST yk280d11.5 comes from this gene; cDNA EST yk286a2.3 comes from this gene"

/codon\_start=1

/protein\_id="CA92434.1"

/db\_xref="PID:e1343689"

/db\_xref="PID:g3873814"

/db\_xref="GI:3873814"

/db\_xref="SPTREMBL:Q17559"

/translation="MAAGIEITFAIRNARFYPSADGSLKIFSLDKMEVEKVAET EAIYGSDAFTEKLNLRNRIEMKORYRTINLNGSTKINGSMGTADPDISMEFAC GERRLLIASPLPCSLGELIVGPDYISQFLRIRFSGNLTSHSPDVLQOLYLISG TGEERTIMLYKSEMLRETKYENNAAFSILPELNFYESSIOAHVNYTPNDDOLVGH CTTTITQLOQGVGHNSYKLMERNKRIHEKICIELEKSELNGSPFFPMENNAKYO LHSADILNSGPNVQSSLSLHHPQSPYDALQIVPPLAYLPNQNPHGILG FAKVQVPGGALQSHCFCLNGTPDPDRVGLGLISARTAVMGQPPAPDFESVI YMSKFAKESRRHGLYFVLIISDGPANALNMKRSIDAIYDASPHMSIIGVMG QORDSHPMNRLEKTLKHSRGLTVKQNSYVSDLESVDYLAIPQMAQKRMFH FDPK"

4953..6585

/gene="C01F6.2"

join(4953..5032,5081..5575,5663..5810,5860..5985,6033..6264,6332..6585)

/gene="C01F6.2"

/codon\_start=1

/protein\_id="CAA92435.1"

/db\_xref="PID:e1343690"

/db\_xref="PID:g3873815"

/db\_xref="GI:3873815"

/db\_xref="SPTREMBL:Q17560"

/translation="MRFADAPHIVYERSEBPWYSGIITAVLIYMSLYTLEA GYRSIRKMLSHMQDQOSTVSTPRCLSTPNMGSKSSPOTRETPDVIYOKPANEPA NCVFIRPVLPKTIPEGTAVPNLNSEEDILLDRDHCIGERKKNLSLEKDKVKEEG KRIKKEIKKSPQMTLDEVQDASSNFNIRPTMAPSPVHTDSTQPSNRKSTGSDS

gene

CDS

NIAAPNSHSEPCOINERVYKQIAGCRNLDEDRPDACACAIISQDSGMDIHGISND FFGKIDQIDQAGDSNILEATSDENLSEMSVTELSMSLAIEPSELETDVQASE TSSHCPEPNAGAKQIVGNEQDADIVDDSHLFESENEGSGSLIVGGQKVFRTTDE KPTTDSVYITGSEFPGMMNINIPMKRNELKMFVCIELPDGMHDYILINYYRFQ" complement(6798..8876)

/gene="C01F6.3"

complement(join(6798..6884,7346..7969,8083..8202,8250..8346,8394..8428,8525..8737,8784..8876))

/gene="C01F6.3"

/note="similar to cytochrome P450; cDNA EST EMBL:CI13310 comes from this gene; cDNA EST EMBL:CI1255 comes from this gene"

/codon\_start=1

/protein\_id="CAA92436.1"

/db\_xref="PID:e1343691"

/db\_xref="PID:g3873816"

/db\_xref="GI:3873816"

/db\_xref="SPTREMBL:Q27466"

/translation="MGVITLAVLASATVIAWLKHLRMQALKHLNOPSRTPIGH GLITRPDEGFENQVIGMGYLPDPRMCLMTGPPCLMLYSGDLVEALFESTKHLNR ILVOKLCCGADERDVLSVIALCTLDICESTMGIAIGAOLAKENNEYVAAVHTLNL ISKRTNPLMNSFIYNYLXGSPILINKENDYKESGLAFLLDILEVNSGOMQETVOA EENTFMLESHDPTSTGLMAVHLLNHPDVQKRYAELDEVGDEDEDITELSSKRY LECALKEALRFPVLIITRELSDQVIGEFIPKGVTFLLNLVYHRDPQAKQPDV FDPDRPHENSIGRVSFAFIPEFSASRNCIGRFAIMEKVIMAHLLRNFNIKAVELM HEPVRKMEIIVPVPIHMKILRRRPIASP"

10227..12369

/gene="fem-3"

join(10227..10297,10396..10740,10786..10864,11236..11583,11666..11756,12137..12369)

/gene="fem-3"

/note="Identical with sex-determining protein FEM-3 (SW:FEW3.CAEEL)"

/codon\_start=1

/protein\_id="CAA92437.1"

/db\_xref="PID:e1343692"

/db\_xref="PID:g3873817"

/db\_xref="GI:3873817"

/db\_xref="SPTREMBL:Q17561"

/translation="MEVDSGSDVDADDETRAKIKIRNRYKRAOMRPFDEYCVTN LTVDDLNPILISGITLQORLTCGATYYDSSLDDQNPWDEFSDRELEITSLITRAGA GTERNDETRFVFORMTKTIIVYCNFMVDLARNRKVOITRELODLIHRDEFERYMY FROFLPNPDPNCTAFSNHTSLHLTYNIPDMQPMNNSQWYNAATGGRVQVONIA AFYPPPEYFNWEDSKYHTTFYVPGCTEFSKFYARF86HGLMPLENETITVLDWLAK ILCEIYVHTTWCIDITGGGGLPRIEHRLAEMENEDIDFDALDPSISRLQISP PEISKSPDVSNGYETIKRKRDIEYQNRLEVYHSDVRIMNYAIDCSKR"

13971..14691

/gene="C01F6.5"

join(13971..14172,14222..14691)

/gene="C01F6.5"

/note="similar to RNA recognition motif (aka RRM, RBD, or RNP domain); cDNA EST EMBL:T00641 comes from this gene; cDNA EST EMBL:T00642 comes from this gene; cDNA EST EMBL:D75737 comes from this gene; cDNA EST EMBL:D75485 comes from this gene; cDNA EST yk269c10.3 comes from this gene; cDNA EST yk269c10.5 comes from this gene"

/codon\_start=1

/protein\_id="CAA92438.1"

/db\_xref="PID:e1343693"

/db\_xref="PID:g3873818"

/db\_xref="GI:3873818"

/db\_xref="SPTREMBL:Q17561"

/translation="MARAIAIDAPISAGIRRTKSPKIRASVGVKKRSPGTAGPSR RPTAOPRROSGIRTVDSISIPANREVRININSNAHYHSDLDQLTAERFIRVSV NFNEHGRPVGIDVTLPKRHADRLQKFRAGVSLDKREKFAITDSANLAKRPEAP RRAPSNGRAQAPRQNVKSGKPTGAGAKKLAKKRPREAOPKKTLEIDAEILDV YMSKA"

24126..25095

/gene="C01F6.6"

join(24126..24194,24453..24532,24780..25095)

/gene="C01F6.6"

/note="similar to PDZ domain (also known as DHR or GLGF).; cDNA EST EMBL:214584 comes from this gene"

/codon\_start=1

/protein\_id="CAA92439.1"

/db\_xref="PID:e1343694"

/db\_xref="PID:g3873819"

/db\_xref="GI:3873819"

/db\_xref="SPTREMBL:Q17562"

/translation="MARAIAIDAPISAGIRRTKSPKIRASVGVKKRSPGTAGPSR RPTAOPRROSGIRTVDSISIPANREVRININSNAHYHSDLDQLTAERFIRVSV NFNEHGRPVGIDVTLPKRHADRLQKFRAGVSLDKREKFAITDSANLAKRPEAP RRAPSNGRAQAPRQNVKSGKPTGAGAKKLAKKRPREAOPKKTLEIDAEILDV YMSKA"

24126..25095

/gene="C01F6.6"

join(24126..24194,24453..24532,24780..25095)

/gene="C01F6.6"

/note="similar to PDZ domain (also known as DHR or GLGF).; cDNA EST EMBL:214584 comes from this gene"

/codon\_start=1

/protein\_id="CAA92440.1"

/db\_xref="PID:e1343695"

/db\_xref="PID:g3873820"

/db\_xref="GI:3873820"

/db\_xref="SPTREMBL:Q17563"

/translation="MARAIAIDAPISAGIRRTKSPKIRASVGVKKRSPGTAGPSR RPTAOPRROSGIRTVDSISIPANREVRININSNAHYHSDLDQLTAERFIRVSV NFNEHGRPVGIDVTLPKRHADRLQKFRAGVSLDKREKFAITDSANLAKRPEAP RRAPSNGRAQAPRQNVKSGKPTGAGAKKLAKKRPREAOPKKTLEIDAEILDV YMSKA"

24126..25095

/gene="C01F6.6"

join(24126..24194,24453..24532,24780..25095)

/gene="C01F6.6"

/note="similar to PDZ domain (also known as DHR or GLGF).; cDNA EST EMBL:214584 comes from this gene"

/codon\_start=1

/protein\_id="CAA92441.1"

/db\_xref="PID:e1343696"

/db\_xref="PID:g3873821"

/db\_xref="GI:3873821"

/db\_xref="SPTREMBL:Q17564"

/translation="MARAIAIDAPISAGIRRTKSPKIRASVGVKKRSPGTAGPSR RPTAOPRROSGIRTVDSISIPANREVRININSNAHYHSDLDQLTAERFIRVSV NFNEHGRPVGIDVTLPKRHADRLQKFRAGVSLDKREKFAITDSANLAKRPEAP RRAPSNGRAQAPRQNVKSGKPTGAGAKKLAKKRPREAOPKKTLEIDAEILDV YMSKA"

24126..25095

/gene="C01F6.6"

join(24126..24194,24453..24532,24780..25095)

/gene="C01F6.6"

/note="similar to PDZ domain (also known as DHR or GLGF).; cDNA EST EMBL:214584 comes from this gene"

/codon\_start=1

/protein\_id="CAA92442.1"

/db\_xref="PID:e1343697"

/db\_xref="PID:g3873822"

/db\_xref="GI:3873822"

/db\_xref="SPTREMBL:Q17565"

/translation="MARAIAIDAPISAGIRRTKSPKIRASVGVKKRSPGTAGPSR RPTAOPRROSGIRTVDSISIPANREVRININSNAHYHSDLDQLTAERFIRVSV NFNEHGRPVGIDVTLPKRHADRLQKFRAGVSLDKREKFAITDSANLAKRPEAP RRAPSNGRAQAPRQNVKSGKPTGAGAKKLAKKRPREAOPKKTLEIDAEILDV YMSKA"

24126..25095

/gene="C01F6.6"

join(24126..24194,24453..24532,24780..25095)

/gene="C01F6.6"

/note="similar to PDZ domain (also known as DHR or GLGF).; cDNA EST EMBL:214584 comes from this gene"



...  
Note: remainder of annotations omitted.

/codon\_start=1  
/protein\_id="CA92439.1"  
/db\_xref="PID:el343694"

Query Match 1.6%; Score 21; DB 21; Length 31281;  
Best Local Similarity 100.0%; Pred. No. 5.82e-01;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1108 TTTAATGATTTGCAGAAA 1128  
|||||  
Oy 1073 TTTAATGATTTGCAGAAA 1093

RESULT 3  
LOCUS ECU73760 2322 bp DNA BCT 05-NOV-1996  
DEFINITION Escherichia coli dehydrin (dhna) gene, complete cds.  
ACCESSION U73760  
VERSION 91658027  
KEYWORDS U73760.1 GI:1658027  
SOURCE  
ORGANISM Escherichia coli.  
Escherichia coli.  
Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia.

REFERENCE 1 (bases 1 to 2322)  
AUTHORS Close, T.J. and Choi, D.-W.  
TITLE Direct Submission  
JOURNALS Submitted (08-OCT-1996) Botany and Plant Sciences, University of  
California, Riverside, CA 92521-0124, USA  
FEATURES  
Source  
Location/Qualifiers  
1..2322  
/organism="Escherichia coli"  
/strain="W3110"  
/db\_xref="taxon:562"  
/map="adjacent to gat operon"  
1106..2230  
/gene="dhna"  
1106..2230  
/gene="dhna"  
/note="similar to cyanodehydrin"  
/codon\_start=1  
/translation="MIARRARRTIHSRYPIGISTYMDIAQLGKADNPLQHR  
MTISDQILYLGHDYVDKVMIDNNRPVLAHNMOTLYNTGRLAGTSLIPDQVE  
HSAGSPANPLFDPKNIVLEAEGNCVASTGYLAVSRRAHRIPLVKNHN  
ETLSYPNTYDQTLVAVQAFNMGAVAGATIGSESRQIEISAFRAHEVGM  
VTVMAYLRNAGFKKGVADYVSADLTGQAHLLATIGADIVKQKMAENNGYAIIN  
GYTDVRYSKLTSENPIDLVRYOLANCLMGAGLINSAGAGGEGDLSDAVRAVIK  
RAGMGILGKARKKSMADVKILINAVQDYILSKITIA"

BASE COUNT 653 a 571 c 589 g 507 t 2 others  
ORIGIN

Query Match 1.5%; Score 20; DB 17; Length 2322;  
Best Local Similarity 100.0%; Pred. No. 3.01e+00;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 90 CCACGCAATATACCTTTT 109  
|||||  
Cp 630 CCACGCAATATACCTTTT 611

RESULT 4  
LOCUS BACSKW01 4043 bp DNA BCT 06-FEB-1999  
DEFINITION Bacillus thuringiensis DNA.  
ACCESSION D86064  
VERSION 91389567  
NID D86064.1 GI:1389567

KEYWORDS  
SOURCE Bacillus thuringiensis (sub-species:sotto, strain:SKW01-10.2-06)  
DNA.  
ORGANISM Bacillus thuringiensis  
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;  
Bacillus.

REFERENCE 1 (bases 1 to 4043)  
AUTHORS Sasaki, J.  
TITLE Direct Submission  
JOURNALS Submitted (17-JUN-1996) to the DDBJ/EMBL/GenBank databases. Jun  
Sasaki, Hokkaido University, Applied life science; N9, W9, Sapporo,  
Hokkaido 060, Japan (E-mail:JSsakai.hines.hokudai.ac.jp,  
Tel:81-11-706-2402)

REFERENCE 2 (bases 1 to 4043)  
AUTHORS Sasaki, J.  
TITLE Bacillus thuringiensis serovar sotto SKW01-10.2-06  
JOURNALS Unpublished (1996)  
FEATURES  
Source  
Location/Qualifiers  
1..4043  
/organism="Bacillus thuringiensis"  
/strain="SKW01-10.2-06"  
/sub-species="sotto"  
/db\_xref="taxon:1428"

BASE COUNT 1430 a 578 c 751 g 1284 t  
ORIGIN

Query Match 1.5%; Score 20; DB 17; Length 4043;  
Best Local Similarity 100.0%; Pred. No. 3.01e+00;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1868 ATACAGATTATATGATTA 1887  
|||||  
Oy 1168 ATACAGATTATATGATTA 1187

RESULT 5  
LOCUS OMTIGHEN 5570 bp DNA VRT 02-JUL-1997  
DEFINITION Oreochromis mossambicus tigh gene.  
ACCESSION Y11732  
NID 92243155  
VERSION Y11732.1 GI:2243155  
KEYWORDS tigh gene; tilapia growth hormone.  
SOURCE Mozambique tilapia.  
ORGANISM Tilapia, mossambica  
Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;  
Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;  
Perciformes; Labroidae; Cichlidae; Tilapia.

REFERENCE 1 (bases 1 to 5570)  
AUTHORS Sekkali, B., Brim, H., Muller, M., Argenton, F., Bortolussi, M.,  
Colombo, L., Belayew, A. and Martial, J.A.  
TITLE A gene encoding tilapia (Oreochromis mossambicus) growth hormone:  
Molecular analysis of the gene and functional test of its promoter  
region by transient expression in vitro and in vivo  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 5570)  
AUTHORS Sekkali, B.  
TITLE Direct Submission  
JOURNALS Submitted (06-MAR-1997) B. Sekkali, Universite de Liege (Ulg),  
Faculte des Sciences, Service de Genie Genetique, Institut de  
Chimie, Bat. B6 Sart Tilman, 4000 Liege, BELGIUM  
FEATURES  
Source  
Location/Qualifiers  
1..5570  
/organism="Tilapia mossambica"  
/db\_xref="taxon:8127"  
/clone="pGEM4.5tigh"  
/clone="pGEM3.5tigh"  
/clone\_id="lambda tigh 1"  
/dev\_stage="adult"  
/tissue\_type="testis"  
1..3679  
/gene="tigh"  
/evidence="experimental"  
1..5149



TATA\_signal  
 /gene="tigh"  
 3631.3637  
 /gene="tigh"  
 join(3680..3754,3850..3983,4405..4521,4645..4788,  
 4861..5007,5087..5345)  
 3680.3754  
 /gene="tigh"  
 /number=1  
 join(3745..3754,3850..3983,4405..4521,4645..4788,  
 4861..5007,5087..5149)  
 /gene="tigh"  
 /codon\_start=1  
 /product="tilapia growth hormone"  
 /protein\_id="CAA72415.1"  
 /db\_xref="PID:e325419"  
 /db\_xref="PID:g2243156"  
 /db\_xref="GI:2243156"  
 /translation="MNSVLLSVYGLGVSSQIINDSORLFIANRYTHLLAORL  
 FSDPSIQTEROLNKIFLODPFNSYIISPIKRETOSSVYKLLISGLVESM  
 EFPKSLSGSSLRNOISPRSELTGILLIRANODERNPDTDLQHAHYGNTYO  
 SLGNSLSRQYELLACFKDKMKVEYLYLVAKRLSPKANCTL"  
 3755..3849  
 /gene="tigh"  
 /number=1  
 3850..3983  
 /gene="tigh"  
 /number=2  
 3984..4404  
 /gene="tigh"  
 /number=2  
 4405..4521  
 /gene="tigh"  
 /number=3  
 4522..4644  
 /gene="tigh"  
 /number=3  
 4645..4788  
 /gene="tigh"  
 /number=4  
 4789..4860  
 /gene="tigh"  
 /number=4  
 4861..5007  
 /gene="tigh"  
 /number=5  
 5008..5086  
 /gene="tigh"  
 /number=5  
 5087..5345  
 /number=6  
 5328..5333  
 POLYA\_signal  
 1766 a 1116 c 1010 g 1678 t  
 BASE COUNT  
 ORIGIN  
 Query Match 1.5%; Score 20; DB 24; Length 5570;  
 Best Local Similarity 100.0%; Pred. No. 3.01e+00;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 2157 AATCATTAATTAACACACA 2176  
 ||||||||||||||||||||  
 Cp 298 AATCATTAATTAACACACA 279  
 ||||||||||||||||||||  
 RESULT 6  
 LOCUS AB011134 5848 bp mRNA PRI 10-APR-1998  
 DEFINITION Homo sapiens mRNA for KIAA0562 protein, complete cds.  
 ACCESSION AB011134  
 NID 93043647  
 VERSION AB011134.1 GI:3043647  
 KEYWORDS KIAA0562 protein.  
 SOURCE Homo sapiens male brain cDNA to mRNA, clone\_11b:pb1uescript11 SK  
 plus clone:H11779.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Ohara,O., Nagase,T. and Ishikawa,K.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-FEB-1998) to the DDBJ/EMBL/GenBank databases. Osamu  
 Ohara, Kazusa DNA Research Institute, DNA Technology, 1532-3 Yana,  
 Kisarazu, Chiba 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,  
 Tel:+81-438-52-3913, Fax:+81-438-52-3914)  
 2 (sites)  
 Nagase,T., Ishikawa,K., Miyajima,N., Tanaka,A., Kotani,H.,  
 Nomura,N. and Ohara,O.  
 TITLE Prediction of the coding sequences of unidentified human genes. IX.  
 The complete sequences of 100 new cDNA clones from brain which can  
 code for large proteins in vitro  
 JOURNAL DNA Res. 5 (1), 31-39 (1998)  
 MEDLINE 98290545  
 FEATURES  
 source Location/Qualifiers  
 1..5848  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="H1179"  
 /clone\_lib="pb1uescript11 SK plus"  
 /sex="male"  
 /tissue\_type="brain"  
 /gene="KIAA0562"  
 314..3091  
 /gene="KIAA0562"  
 314..3091  
 /codon\_start=1  
 /product="KIAA0562 protein"  
 /protein\_id="BAA25488.1"  
 /db\_xref="PID:d1026418"  
 /db\_xref="PID:g3043648"  
 /db\_xref="GI:3043648"  
 /translation="MPKIGFYVYSSSEHEGFSARELMITAPYSGRSRPFQCFQ  
 EIVLQWVERKIRKIQILAHQYMTSSKIEFTISSLPYFAPYQERRRIGVSLCD  
 NEKTCCKARLEKSYVDAVGOFLKIFQNHVKNYININOVLAVINIGDPADFSDE  
 SNTASREKLIIDHYGHNSDEPDALGETYARKDYSIPDLDFADMDQDEVAQIIRKLD  
 EKRREAVOKERYDYARKLKLOAIADLOKGEKREYRCARCAVEDKEDVLAKEKKQK  
 EOYAEVYEOLEISILDAELMRBPDPLOPLARSGSCPCOKMPSPOLIEBECTEN  
 QPAEPFLOEKSSYSITISPOHSAVDPLPATDPKINAESLPIDENPLPAIRKHG  
 EAVEPEMSNADISDARRGMLGEPEPTELEAEASSAIDVLETTVAEAYCKTWY  
 REDALLASKLMEPVGTPKEDKNIRASVPLVRAIKDIVSVFQASLKLMI  
 TOYIPKHKLSKLETAHCYERTIPVLTPTGSSARLVYTAANFIOEMALFKEVSLQI  
 IPSYLVOLKANSVHLAMSGMLARLKLGLGSSGFTIDNWKKEFVSALDEHRYE  
 VRETAVRITIDMYRHOHASTILEYLPDPDSNTRRNILYKTIFEGPAKIDGRATDAEMRA  
 RKKATEAEKOKKEIKALOGLALEKIOAEVOKESDVKRPKNODIQGKAPAE  
 ALGIPDEHYLDNLICIFGERSSEFTLEBGLHWYKHCHMLRCHCKOVHISLSTEH  
 LITBEDKKDGFCKCYRCSEAVFKEELPHIKHKDCNPAKPERKANRCPLCHENFSPGE  
 EAMKRLHMGPAAGCTMNLKTHILQKAPALQGKSSAVAAAGPLSGSKASKIPPTKGL  
 SKSSRYVAKR"  
 BASE COUNT 1582 a 1387 c 1417 g 1462 t  
 ORIGIN  
 Query Match 1.5%; Score 20; DB 29; Length 5848;  
 Best Local Similarity 100.0%; Pred. No. 3.01e+00;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 4965 ATAAAAAGACTTCATTTTA 4984  
 ||||||||||||||||||||  
 Oy 398 ATAAAAAGACTTCATTTTA 417  
 ||||||||||||||||||||  
 RESULT 7  
 LOCUS AE000299 10036 bp DNA BCT 12-NOV-1998  
 DEFINITION Escherichia coli K-12 MG1655 section 189 of 400 of the complete  
 genome.  
 ACCESSION AE000299 U00096  
 NID 91788413  
 VERSION AE000299.1 GI:1788413  
 KEYWORDS Escherichia coli.  
 SOURCE



```

ORGANISM      Escherichia coli
REFERENCE      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
AUTHORS        Escherichia.
                1 (bases 1 to 10036)
                Blatter,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,Y.,
                Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
                Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
                Mau,B. and Shao,Y.
                The complete genome sequence of Escherichia coli K-12
                Science 277 (5331), 1453-1474 (1997)
MEDLINE        97426617
REFERENCE      2 (bases 1 to 10036)
AUTHORS        Blatter,F.R.
TITLE          Direct Submission
JOURNAL        Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
                University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
                Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
                608-263-7459
                3 (bases 1 to 10036)
                Blatter,F.R.
                Direct Submission
                Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
                University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
                Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
                608-263-7459
                4 (bases 1 to 10036)
                Plunkett,G. III.
                Direct Submission
                Submitted (13-OCT-1998) Laboratory of Genetics, University of
                Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
                This sequence was determined by the E. coli Genome Project at the
                University of Wisconsin-Madison (Frederick R. Blattner, director).
                Supported by NIH grants HG00301 and HG01428 (from the Human Genome
                Project and NCHGR). The entire sequence was independently
                determined from E. coli K-12 strain MG1655. Predicted open reading
                frames were determined using Genemark software, kindly supplied by
                Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
                30332 [e-mail: markborodov@gatech.edu]. Open reading frames that
                have been correlated with genetic loci are being annotated with CG
                Site Nos., unique ID nos. for the genes in the E. coli Genetic
                Stock Center (CGSC) database at Yale University, kindly supplied by
                Mary Berrlyn. A public version of the database is accessible
                (http://cgsc.biology.yale.edu). Annotation of the genome is an
                ongoing task whose goal is to make the genome sequence more useful
                by correlating it with other data. Comments to the authors are
                appreciated. Updated information will be available at the E. coli
                Genome Project's World Wide Web site
                (http://www.genetics.wisc.edu). *** The E. coli K-12 sequence and
                its annotations are periodically updated; this is version M54. No
                sequence changes. Annotation updates: updated gene identifications
                and products; all new functional assignments courtesy of Monica
                Riley; added promoters, protein binding sites, and repeated
                sequences described in reference 1. The unique numeric identifiers
                beginning with a lowercase 'b' assigned to each gene (protein- or
                RNA-encoding) are now designated as gene synonyms instead of
                labels. This should allow them to be searched for in Entrez as gene
                names.
FEATURES
source         1. 10036
                /organism="Escherichia coli"
                /strain="K-12"
                /sub_strain="MG1655"
                /db_xref="taxon:562"
                77. 153
                /note="RFP (repetitive extragenic palindromic) element;
                contains 2 RFP sequences"
                complement(181..1305)
                /gene="b2097"
                /gene="b2097"
                complement(181..1305)
                /gene="b2097"
                /function="orf; unknown"
                /note="f374; This 374 aa ORF is 30 pct identical (9 gaps)
                to 102 residues of an approx. 512 aa protein FLIC_SALMU
                RE
REFERENCE      1 (bases 1 to 10036)
AUTHORS        Blatter,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,Y.,
                Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
                Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
                Mau,B. and Shao,Y.
                The complete genome sequence of Escherichia coli K-12
                Science 277 (5331), 1453-1474 (1997)
MEDLINE        97426617
REFERENCE      2 (bases 1 to 10036)
AUTHORS        Blatter,F.R.
TITLE          Direct Submission
JOURNAL        Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
                University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
                Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
                608-263-7459
                3 (bases 1 to 10036)
                Blatter,F.R.
                Direct Submission
                Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
                University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
                Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
                608-263-7459
                4 (bases 1 to 10036)
                Plunkett,G. III.
                Direct Submission
                Submitted (13-OCT-1998) Laboratory of Genetics, University of
                Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
                This sequence was determined by the E. coli Genome Project at the
                University of Wisconsin-Madison (Frederick R. Blattner, director).
                Supported by NIH grants HG00301 and HG01428 (from the Human Genome
                Project and NCHGR). The entire sequence was independently
                determined from E. coli K-12 strain MG1655. Predicted open reading
                frames were determined using Genemark software, kindly supplied by
                Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
                30332 [e-mail: markborodov@gatech.edu]. Open reading frames that
                have been correlated with genetic loci are being annotated with CG
                Site Nos., unique ID nos. for the genes in the E. coli Genetic
                Stock Center (CGSC) database at Yale University, kindly supplied by
                Mary Berrlyn. A public version of the database is accessible
                (http://cgsc.biology.yale.edu). Annotation of the genome is an
                ongoing task whose goal is to make the genome sequence more useful
                by correlating it with other data. Comments to the authors are
                appreciated. Updated information will be available at the E. coli
                Genome Project's World Wide Web site
                (http://www.genetics.wisc.edu). *** The E. coli K-12 sequence and
                its annotations are periodically updated; this is version M54. No
                sequence changes. Annotation updates: updated gene identifications
                and products; all new functional assignments courtesy of Monica
                Riley; added promoters, protein binding sites, and repeated
                sequences described in reference 1. The unique numeric identifiers
                beginning with a lowercase 'b' assigned to each gene (protein- or
                RNA-encoding) are now designated as gene synonyms instead of
                labels. This should allow them to be searched for in Entrez as gene
                names.
FEATURES
source         1. 10036
                /organism="Escherichia coli"
                /strain="K-12"
                /sub_strain="MG1655"
                /db_xref="taxon:562"
                77. 153
                /note="RFP (repetitive extragenic palindromic) element;
                contains 2 RFP sequences"
                complement(181..1305)
                /gene="b2097"
                /gene="b2097"
                complement(181..1305)
                /gene="b2097"
                /function="orf; unknown"
                /note="f374; This 374 aa ORF is 30 pct identical (9 gaps)
                to 102 residues of an approx. 512 aa protein FLIC_SALMU
                RE
                3765..4730
                /gene="b2100"
                /gene="b2100"
                complement(181..1305)
                /gene="b2100"
                /function="putative enzyme; Not classified"
                /note="o321; This 321 aa ORF is 26 pct identical (15 gaps)
                to 297 residues of an approx. 312 aa protein RBSK_ECOLI
                SW: P05054"
                /codon_start=1
                SW: P06177"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC75158.1"
                /db_xref="PID:g1788414"
                /db_xref="GI:1788414"
                /translation="MARRRRRTTISRYPDIGYSIVMTDIAQLCKADNLLQHRG
                MTPISDLYLPGHDYVDVDMVNNRPVAVRNQIYNTGRLAGTGLSILPYDQVE
                HSAGSPFANPLVEPKNIYELAEACNCAVSTGYLAVSRRAHRIFFLVKLNH
                ETLSPYPTDPTLYASVEQAFNMGAVAVGATITFGESESRROELTSAFERAHELM
                VTYIMATLRNSAKKQSDVDYHVSADLTGQANHATIGATIVYQKNAENNGKKAITY
                GTYDDRYTSKLTISENPLDLYRIQANCYWGRAGLINSGGAGGETDLSDAVRAIVLK
                RAGGMGLILGRKAKKSMADGKRLINAVDPIYDSKITTA"
                complement(1308..1336)
                /note="factor Sigma70; predicted +1 start at 2176652"
                1405..1434
                /note="factor Sigma70; predicted +1 start at 2176792"
                1420..1447
                /note="factor Sigma70; predicted +1 start at 2176805"
                1490..1767
                /gene="b2098"
                /gene="vegT"
                1430..2767
                /gene="vegT"
                /function="putative transport; Not classified"
                /note="o425; This 425 aa ORF is 32 pct identical (30 gaps)
                to 401 residues of an approx. 424 aa protein NUPG_ECOLI
                SW: P09452"
                /codon_start=1
                /transl_table=11
                /product="putative nucleoside permease protein"
                /protein_id="AAC75159.1"
                /db_xref="PID:g1788415"
                /db_xref="GI:1788415"
                /translation="MKTATLSFMMFVEMFINGAMPVPLMLNLSKSGFSGAEIGNSYA
                CHTAATLSPILVGSITDRFSKQKVALVMTFGALMTFAAQITFAGFIPLLATS
                LTMPTIALNLSIAFANVPVEDPFRIRVGTIGWASGLACGFLPQIIGVADISPT
                NIPILLTSGSSALGVAFELPPTPKSTGMDIKWLGIDALILRDKNLEFFVCS
                FLFAMPATFYIRFANGYLTREVGMKNATGWTLCOFSEIFMLALFFETKRGJAKVLL
                LGIVTAIRYGEFTFISADYFTYVALLFGLILHGVSYDYYVYAVYVYVYKRAVDHNR
                TAAGLTLLCCGFGSLGTRGCGVMEKRFAPCEPNGLTFMWSGMTGAVMIATL
                AVLFMTFRSDNEITRAIKVDRDIALTOGEVR"
                2764..3768
                /gene="b2099"
                /gene="b2099"
                2764..3768
                /gene="b2099"
                /function="orf; Unknown"
                /note="o334; This 334 aa ORF is 27 pct identical (16 gaps)
                to 188 residues of an approx. 296 aa protein DRAG_RHOURU
                SW: P14300"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC75160.1"
                /db_xref="PID:g1788416"
                /db_xref="GI:1788416"
                /translation="MKTERTLIGALYGOALGDAMGMPSELPRSRVKAHFGMIDREFLPG
                PKENNAACYFNRAPFETDDTSALCIDALLERKGDIDPDIGIITIMDALRPAFKN
                VLPSTKTIANAIIRDGKPVAELENNGYTNKAARVPSGLGLRLARDVSDTIDVVALAS
                SPYHKSDLAVAGAVVIAWATSRVIDSGSAIVDSPTSRHAQOKRTTFSSALAR
                LELALKIVRNADGETESASDOLYOVVAGSTISISVPALAVLEAOTDPRCAVILAN
                LGSDTDLTGAMATATGALHGVNAIDPALKAEDDAVAVQDFNRYATALAKYRQOREAV
                "
                3765..4730
                /gene="b2100"
                /gene="b2100"
                complement(181..1305)
                /gene="b2100"
                /function="putative enzyme; Not classified"
                /note="o321; This 321 aa ORF is 26 pct identical (15 gaps)
                to 297 residues of an approx. 312 aa protein RBSK_ECOLI
                SW: P05054"
                /codon_start=1

```



```

/transl_table=11
/product="putative kinase"
/protein_id="AAC75161.1"
/db_xref="P010:91788417"
/db_xref="GI:1:1788417"
/transl_table="MSSAARLTLPELTTROSVMVGAVIDADAYALPMWRCDEIHEI
LKSQVNVGGCALUNIAVALRLILTEGNNALPLGGQVWAMIMINRAKGLSLIDNME
GDNMNCALVADPEERDEPSESEVENONKROWALRYLTPAGSLTFSQGLASPGSEI
LVNMLLEADYTPRIDPGRICGIPDALLARLITACRPLYSLNROAEIYLAERAPLSAE
ITTLTGKQMOEFAFPLPILRIDKSGAMVFSNDASGCLPAPIGVVDITIGASGSHGAVT
AGLASGPELADVATLGNAAVSWVGHRRGDCAPTREELLAKHNV"
complement(4704..5450)
/gene="yegw"
complement(4704..5450)
/gene="yegw"
/function="putative regulator; Not classified"
/note="f248; This 248 aa ORF is 30 pct identical (14 gaps)
to 217 residues of an approx. 272 aa protein YHR_0COI1
SM: P45544"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAC75162.1"
/db_xref="P010:91788418"
/db_xref="GI:1:1788418"
/transl_table="MEDAHQQLIAQINERILTAADNTPLYIKAEIYKKNVSGVLEHNG
NLTGEEHDSLSQPLTSMNRYIVTKRMALGEEGVATRSRGITGINNIFYSILKEKRSFS
QNVVALRKRKDPVTLMVNNRKYVCAEVAQDLAEVAGSDVFLRLIRYVDEAEVSEES
VPARLHIDVAVAGISLTDYRSQHIYPORTRSKVSARMPDAEFOSHIOLDSKIPVLVY
KOVALDOOORPIEISIHCSGSDLYFVCEE"
complement(5449..5526)
/note="factor Sigma20; predicted +1 start at 2180843"
complement(5502..6529)
/gene="b2102"
/note="b2102"
/gene="yegw"
complement(5502..6529)
/gene="yegw"
/function="orf; Unknown"
/note="f275; This 275 aa ORF is 36 pct identical (21 gaps)
to 206 residues of an approx. 296 aa protein LYCM_STR6L
SM: P25310"
/codon_start=1
/transl_table=11
/product="orf; hypothetical protein"
/protein_id="AAC75163.1"

```

Query Match	1.5%:	Score 20:	DB 18:	Length 10036:
Best Local Similarity 100.0%:		Pred. No.	3.01e+00:	
Matches	20; Conservative	0;	Mismatches 0;	Indels 0; Gaps 0;
Db 2303	AAAAAGCATATATGCTTGG	2322		
Oy 611	AAAAAGCATATATGCTTGG	630		

Note: remainder of annotations omitted.

RESULT	8				
LOCUS	u67574	11978 bp	DNA	BCT	20-MAY-1998
DEFINITION	Methanococcus jannaschii section 116 of 150 of the complete genome.				
ACCESSION	U67574	L77117			
NID	91591978				
VERSION	067574.1	GI:1591978			
KEYWORDS					
SOURCE	.				
ORGANISM	Methanococcus jannaschii.				
	Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;				
	Methanococcus.				
	1 (bases 1 to 11978)				
REFERENCE	Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D.,				
AUTHORS	Sutton,G.G., Blake,J.A., Fitzgerald,L.M., Clayton,R.A.,				
	Gocayne,J.D., Kellavagge,A.R., Dougherty,B.A., Toml,J., Adams,M.D.,				

TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS

2 (bases 1 to 11978)  
Bult,C.U., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D.,  
Sutton,G.G., Blake,J.A., Fitzgerald,L.M., Clayton,R.A.,  
Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J., Adams,M.D.,  
Reich,C.I., Overbeck,R., Kirkness,E.F., Weinstock,K.G.,  
Merrick,J.M., Glocke,A., Scott,J.D., Geoghagen,N.S., Weidman,J.F.,  
Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A.,  
Roberts,K.M., Kaine,B.B., Borodovsky,M., Klenk,H.P., Fraser,C.M.,  
Smith,H.O., Moese,C.R. and Venter,J.C.

Complete genome sequence of the methanogenic archaeon,  
*Methanococcus jannaschii*  
Science 273 (5278), 1058-1073 (1996)  
96337999

TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS

2 (bases 1 to 11978)  
Bult,C.U., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D.,  
Sutton,G.G., Blake,J.A., Fitzgerald,L.M., Clayton,R.A.,  
Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J., Adams,M.D.,  
Reich,C.I., Overbeck,R., Kirkness,E.F., Weinstock,K.G.,  
Merrick,J.M., Glocke,A., Scott,J.D., Geoghagen,N.S., Weidman,J.F.,  
Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A.,  
Roberts,K.M., Kaine,B.B., Borodovsky,M., Klenk,H.P., Fraser,C.M.,  
Smith,H.O., Moese,C.R. and Venter,J.C.

Direct Submission  
Submitted (27-AUG-1996) The Institute for Genomic Research, 9712  
Medical Center Dr. Rockville, MD 20850, USA  
On Oct 3, 1996 this sequence version replaced gi:1565379.  
Location/Qualifiers  
I..11978

gene 145. .840  
/gene="MJ1337"  
145. .840  
/gene="MJ1337"  
/note="hypothetical protein; identified by Genemark;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="Jannaschii predicted coding region MJ1337"  
/protein\_id="AB99358.1"  
/db\_xref="PID:G1592330"  
/db\_xref="GI:1592330"  
/translation="MIGAVLKTLEYFDEKDVKITTTGDIGEGDSLKLYDALKEIDDD  
LYIIVYIKPKIKREINFEPSKRIADAGMAAKAANTGDFEYLFDPVGLAFAD  
KSHVAYRGFSEIDDNVEPLIRDYKLTMPKMYVKGEDYIIREKIIIDIKKEE  
KIKAECCIGGCDTLTEIVSLISVDFTEALSLGCKINKKLGIANVANTQTITE  
INATKALNTLTK"  
917. .1978  
/gene="MJ1338"  
917. .1978  
/gene="MJ1338"  
/note="similar to SP:P32440 GB:U19363 GB:U19364 GB:X55547  
PID:48909 percent identity: 29.09; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="H(2)-dependent methylenetetrahydromethanopterin  
dehydrogenase related protein"  
/protein\_id="AB99348.1"  
/db\_xref="PID:G1591980"  
/db\_xref="GI:1591980"  
/translation="MRNRKIKVDNMKRSVYGAGNONLYINKLNPENKFGGEPPYGG  
RAIFAFAGHGVVLAEPNRKIMSDLDKRYEDAGKVYSDVEAKGELHVFEPF  
GATPRKIKTIIEHPNNAVINCTVSPVVLVLSLEPILTRKRDVGISMHBAAV  
GPGHGVIGKTTGDKRELATEBOIKRAVLAKSAGKEAVVPADVSVAADGSLV  
TVALSGVDYTYVGRKTIINAPKKHIEQVIMTLQTMASLVETSGISEMVAALNP  
ELLLSASGKILDKDQDLALETLLQNLDETLEAVEKAEIKFTTLVAOSLVKEIKTLL  
GGAAEGAKIRKARLFEH"  
2010. .2474  
/gene="MJ1339"  
2010. .2474  
/gene="MJ1339"  
/note="similar to SP:P04766 GB:X04399 PID:39954 percent  
identity: 32.65; identified by sequence similarity;  
putative"







TITLE Smith, H.O. and Venter, J.C.  
Whole-genome random sequencing and assembly of Haemophilus  
influenzae Rd  
JOURNAL Science 269 (5223), 496-512 (1995)  
MEDLINE 95350630  
REFERENCE  
AUTHORS 2 (bases 1 to 12188)  
Tatusov, R.L., Mushegian, A.R., Bork, P., Brown, N.P., Hayes, W.S.,  
Borodovsky, M., Rued, R.E. and Koonin, E.V.  
Metabolism and evolution of Haemophilus influenzae deduced from a  
whole-genome comparison with Escherichia coli  
JOURNAL Curr. Biol. 6 (3), 279-291 (1996)  
MEDLINE 96398784  
REFERENCE  
AUTHORS 3 (bases 1 to 12188)  
White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.  
Direct Submission  
JOURNAL Submitted (25-JUL-1995) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
4 (bases 1 to 12188)  
White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.  
Direct Submission  
JOURNAL Submitted (27-SEP-1997) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
The H. influenzae sequence has been updated by R. Fleischmann. New  
database matches have been assigned, product names have been  
improved, and a number of frame shifts have been corrected. We  
gratefully acknowledge the work of Tatusov et. al. We have  
incorporated their annotation into the /notes fields of the  
corresponding H. influenzae genes  
5 (bases 1 to 12188)  
White, O., Clayton, R.A., Kerlavage, A.R., Fleischmann, R.D.,  
Peterson, J., Hickey, E., Dodson, R. and Gwinn, M.  
Direct Submission  
JOURNAL Submitted (28-MAY-1998) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
The whole genome was shifted by 588 nucleotides for a new start  
on Jun 12, 1999 this sequence version replaced gi:1574746.  
Location/Qualifiers  
1. 12188  
/organism="Haemophilus influenzae Rd"  
/db\_xref="taxon:71421"  
288. 674  
/gene="HI1288"  
288. 674  
/note="similar to SP:P09170 GB:X13270 PID:42218 PID:42222  
PID:606107 percent identity: 67.44; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="ribosome binding factor A (rbfa)"  
/protein\_id="AAC22937.1"  
/db\_xref="PID:g1574747"  
/db\_xref="GI:1574747"  
/translation="MARFKRSQRAOEIOKEAVLQIEVKDPRIGMTVSDVEVSS  
DLSYKIVTFLEHDEMAIEQMGLEKASPIRSLKAMKRLRTVPEIRFLPYDSL  
VGMRMSNLVTNVNREDEKHYEESN"  
674. 1612  
/gene="HI1289"  
674. 1612  
/note="similar to SP:P09171 GB:X13270 PID:42219 PID:42223  
PID:606106 percent identity: 58.44; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="tRNA pseudouridine 55 synthase (trub)"  
/protein\_id="AAC22938.1"  
/db\_xref="PID:g1574748"  
/db\_xref="GI:1574748"  
/translation="MSRPRKRWDRVDVGVFLDQPGSSNDIMQYKRLFOANKAGT  
GALDPLATGMLPCLGSEAKRFESQFLDADKRYLYTKLGERDTSDEAGGVETREYN  
LEQQITLAEORGDILQVPIFSAKHNKGRPLVYARQGITVEREAPRTITFELNF  
IEYNAPFLTLEVHCSKGIYIRLVLDLGEVLCGAAVTMLRFAVADVPVAKEMPINE

gene  
CDS  
1672. 2805  
/gene="HI1290"  
1672. 2805  
/note="similar to SP:002287 GB:M74135 GB:X60420 PID:415010  
PID:43345 percent identity: 58.86; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="chorismate mutase / prephenate dehydrogenase  
(tyra)"  
/protein\_id="AAC22939.1"  
/db\_xref="PID:g1574749"  
/db\_xref="GI:1574749"  
/translation="MSFMEALKDSREISDIRELIQFAKRLIEVSOYGYKHQHG  
PIYABERLAMIQARLEAEKAGISADLIEDVLRFRMSYANENOFGRKINSIHK  
IVIVGYGKLGGLFARYLAKASGPISTIDREDAVAESTLANADVAIVSVPTINLET  
IERLKPYLENNLADLTSVKREPLAKMLEVHTGAVLGHPMEGADISMAKQVYVRC  
DERFEREYMLEEQIOWGAKIYQTNATEHDHNTYIQALRHFSTFANGLSKOPIN  
LANLALSPYRLLEAMIGRLFAQDAELIADIIDKSENLAIVFTLKQTYDEALTFE  
ENNDROGFTDAHKYRDMFGDYSRQFLKESRQLQOANDLKQG"  
complement(2838. 3677)  
/gene="HI1291"  
complement(2838. 3677)  
/note="similar to PID:882689 GB:U00096 SP:Q46920  
PID:1789158 percent identity: 61.59; identified by  
sequence similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAC22940.1"  
/db\_xref="PID:g1574750"  
/db\_xref="GI:1574750"  
/translation="MNYQDNLKSLKLGQTEYASQYDFTLQPVPRALNRDGLGITO  
NOPFTIGADIMTAYEISMLNEKGLPOVAIADYIDYOSQNLIESKFKYLNSFNQSK  
FADFNVAQOTMORDSECOGQVYKRLNPMAYDSOKIDHLOGDIDEDDITSTSEF  
NANMLKDCVSDIEVEKLYSHLKSNCILITNOPMGTLAHIVHGKIKNEKILRYVVS  
FROHNEFHQCVERTFCDLMAHYAKREKLYVARYTRRGSLDINPRSNENPENLRL  
AKQ"  
complement(3804. 4589)  
/gene="HI1292"  
complement(3804. 4589)  
/note="hypothetical protein: identified by Genemark;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="H. influenzae predicted coding region HI1292"  
/protein\_id="AAC22948.1"  
/db\_xref="PID:g1574751"  
/db\_xref="GI:1574751"  
/translation="MYLLETYRLTALENNIESQRLNNAVVDQRYNGQILGRIPL  
VLAEEAGQGFAMRVICPEODSLFPQNNAAEVNRLQOEKKGVLFGDGLGDFENS  
DOTAEVASPAQMOVLVYTHLOSCSPHSGENFAPRLQKKNPHLTODLQIKWQMO  
ACDLOMNGAVLEQOSIAEISDHQSTLSHGXYLQAEIEKEKGIPTYYLYLVAGQSL  
KSEKSRCCPSGANNALKAIDITFTFKDCTRLVSNLSWNL"  
complement(4580. 4960)  
/gene="HI1293"  
complement(4580. 4960)  
/note="similar to PID:882706 GB:U00096 SP:Q46926  
PID:1789176 percent identity: 36.89; identified by  
sequence similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAC22947.1"  
/db\_xref="PID:g1574752"  
/db\_xref="GI:1574752"  
/translation="MIEQLKQAKNMDRYRLIIQAGKNLPRPSDNLQMQPITCEA







```

gene
complement(1007. .1291)
/gene="gacB"
CDS
complement(1007. .1291)
/gene="gacB"
/Note="ORF_ID:0358#3; similar to [SwissProt Accession
Number P37188]"
/codon_start=1
/transl_table=11
/product="PTS system, Galactitol-specific IIB component
(EIIB-GAT) (Galacticol- permease IIB component)
(Phosphotransferase enzyme II, B component) (EC
2.7.1.69)."
/protein_id="BAA15960.1"
/db_xref="PID:d1016685"
/db_xref="PID:g1736816"
/db_xref="GI:1736816"
/translation="MKRKIVAGGAVATSTMAAEIKELCONHNIPVELIQCVRNVI
ETVMDGVHLICTAKVDRSGDIPVHGMPFISGIGIEALQNKILITLQ"
complement(1322. .2973)
/gene="gata"
/Note="ORF_ID:0358#4; similar to [SwissProt Accession
Number P37187]; start codon is not identified yet"
/codon_start=1
/transl_table=11
/product="PTS system, Galactitol-specific IIA component
(EIIA-GAT) (Galacticol- permease IIA component)
(Phosphotransferase enzyme II, A component) (EC
2.7.1.69)."
/protein_id="BAA15961.1"
/db_xref="PID:d1016686"
/db_xref="PID:g1736817"
/db_xref="GI:1736817"
/translation="IMNPPVLMLOMTNMISYQGLVTRFPKSAIYLLRPYKHFQQA
DDNDVAVSLVIALIVENPQQLKLCRLGKLGQDPVETLITLLETQLEKFTKIV
LDSDE"
1605. .2621
/gene="IS5"
1605. .2621
/gene="IS5"
/Note="ORF_ID:0359#1; similar to [PIR Accession Number
A91483]"
/codon_start=1
/transl_table=11
/protein_id="BAA15962.1"
/db_xref="PID:d1016687"
/db_xref="PID:g1736818"
/db_xref="GI:1736818"
/translation="MEVIMSHRTGFTIMSHQUTFEADSEFSSKRROTREKIFLSMEQIL
PMQNVIEVEPEFYKAGNGRPPYLEMLRIHOMQWNYNSDAMDALYEISMLE
ARLSALPDRITINPRLLEOHOLAROLFETIRMYLAEAQVMMTCGLINDATIE
ABSTKNKQQRDPREKHOTKKQNMHFGKAHIGVAKSGLTSLVTTANENDLQNL
GNIHSEQVSADAGIQGAPQRELEADVDMVLAERPEKVFLLQHPKRNKTAINI
EYKVASIRAVEHPRIKQFGFVARVKGLKNDNQLAMFTLANLEFRADQMIRQW
ERSH"
complement(1917. .2273)
/gene="IS5"
/Note="ORF_ID:0359#2; similar to [PIR Accession Number
B91483]"
/codon_start=1
/transl_table=11
/protein_id="BAA15963.1"
/db_xref="PID:d1016688"
/db_xref="PID:g1736819"
/db_xref="GI:1736819"
/translation="MOQITQLIEVMLVGRGQDAVQATLGLDITNVGLAKYPLIPL
GLMLRIALLFVLGRAGCLINDGIDQALSHHDACGQPAIDGLQLAGQMLLQOV
AEIHDGAVROGAIQ"
complement(2659. .2973)
/gene="gata"
/Note="ORF_ID:0359#4; similar to [SwissProt Accession

```

```

Number P37187]"
/codon_start=1
/transl_table=11
/product="PTS system, Galactitol-specific IIA component
(EIIA-GAT) (Galacticol- permease IIA component)
(Phosphotransferase enzyme II, A component) (EC
2.7.1.69)."
/protein_id="BAA15964.1"
/db_xref="PID:d1016689"
/db_xref="PID:g1736820"
/db_xref="GI:1736820"
/translation="MTNLFVRSGLSVDSSEVLTIGNEMKANGVHDTPQALLARE
AEPFTGIMLQHAIALPHCEALHAKEGANKRGNSSRLTDSFHFMEPIFSPVALNQ
PI"
complement(2983. .4245)
/gene="gatz"
complement(2983. .4245)
/gene="gatz"
/Note="ORF_ID:0359#5; similar to [SwissProt Accession
Number P37191]"
/codon_start=1
/transl_table=11
/product="Putative tagatose 6-phosphate kinase Gatz (EC
2.7.1.-)."
/protein_id="BAA15965.1"
/db_xref="PID:d1016690"
/db_xref="PID:g1736821"
/db_xref="GI:1736821"

```

Query Match 1.5%; Score 20; DB 17; Length 17913;  
 Best Local Similarity 100.0%; Pred. No. 3.01e+00;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7557 AAAAGGTATTATTCCTGG 7576  
 611 AAAAGGTATTATTCCTGG 630

```

RESULT 11
LOCUS CELC17H12 47045 bp DNA INV 11-FEB-1998
DEFINITION Caenorhabditis elegans cosmid C17H12.
ACCESSION AF045642
VERSION 52854160
KEYWORDS AF045642.1 GI:2854160
SOURCE Caenorhabditis elegans strain-Bristol N2.
ORGANISM Caenorhabditis elegans
Eukaryote; Metazoa; Nematoda; Secernentea; Rhabditi; Rhabditiidae;
Rhabditina; Rhabditidae; Rhabditidae; Pelodertinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 47045)
WILSON,R., Almscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Crahan,M., Dear,S., Du,Z., Durbin,R., Favell,A.,
Fullon,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., McMurtry,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Riken,L., Roop,A.,
Saunders,D., Showkeen,R., Smailson,N., Smith,A., Sonhammer,E.,
Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
Wilkinson-Sprat,J. and Wohldman,P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL NATURE 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 47045)
AUTHORS Geisel,C.
TITLE The sequence of C. elegans cosmid C17H12
JOURNAL Unpublished (1998)
REFERENCE 3 (bases 1 to 47045)
AUTHORS Waterston,R.

```







25635..25901,26165..26449)  
 /gene="C17H12.4"  
 /note="Similar to carboxylesterase; coded for by C.  
 elegans CDNA YK299g11.5"  
 /codon\_start=1  
 /protein\_id="AAC002583.1"  
 /db\_xref="PID:G2834165"  
 /db\_xref="GI:2854165"  
 /translation="MLPLGLNNLLFFLFFLSITKSAIVSTSLQIDQDQISF  
 HLEHFLPLAKPRIGLRFQKPEPEKMTGVNAKAYGPACISNSYSTSPQKIMEDC  
 LHINFTSPDKCKSKDCAVYVEHGGGLMYDSAVNENPVLGTFVNDVITAIAPR  
 LQIFSHFVEDDSIAPTNLALDILGVFVNEHNEHREGNNOKNTIKHSGSIVN  
 ILTFSSKINHDLSTFQOSTCMSGSDHFELELQIOKTFNFAKHAGCTVPSILKMTY  
 ROSDLYRMKCLQNDKDLLELRYQRLDEDEGPTGYLVQREPLQIEIPQEMKSPK  
 IPMLGCTREEMDSHPAPKPGIQAALGFENMEYKFRKMDDEVFEFNSDEQAI  
 MVQVRVYNLNLKSNPAYLYETYPKHARHTDLEFLMGVRFEDENETELGKYVE  
 TLENNAPKDPKREGPELANMESYEVYMDGEPGERPOMTKFEKILINWMDMV  
 EYDKHTELEKKNKISIRKPYVFPFESSKTYHOYSWTISILILFTILFTGLFGKCC  
 VGNDRNLTYIQDGSNTPYGHVKNP"  
 complement(26876..29985)  
 /gene="C17H12.12"  
 complement(26876..27060,27174..28035,28087..28201,  
 complement(join(26876..27060,27174..28035,28087..28201,

Note: remainder of annotations omitted.

Query Match 1.5%; Score 20; DB 21; Length 47045;  
 Best Local Similarity 100.0%; Pred. No. 3.01e+00;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 20569 AACATCTAATTCAATTGAT 20588  
 ||||||||||||||||||  
 Cp 282 AACATCTAATTCAATTGAT 263

RESULT 12  
 LOCUS AC0005285 209071 bp DNA INV 15-JUL-1998  
 DEFINITION Drosophila melanogaster DNA sequence (Pls DS00121 (D128), DS05470  
 (D270), and DS00108 (D120)), complete sequence.  
 AC0005285 AC0004267 AC0002637 AC0003139 AC0003717 AC0003718  
 AC0003140 AC0003141 AC0002639 AC0003142 AC0003720 AC0003143  
 AC0004568 AC0004269 AC0002586 AC0002587 AC0003130 AC0002588 AC0002589  
 AC0002590 AC0003131 AC0003709  
 93320127  
 AC0005285.1 GI:3320127  
 HTG.  
 ORGANISM Drosophila melanogaster (Subclones in sac from P1 clones DS00121  
 (D128), DS05470 (D270), and DS00108 (D120)) DNA.  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;  
 Drosophilidae; Drosophila.  
 1 (bases 1 to 209071)  
 Celiker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,  
 Switskas,R.R., Harris,N.L., Agapayani,A., Arcinas,T.T., Baxter,E.,  
 Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,  
 Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
 Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,  
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
 Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,  
 Yee,A., Zhang,R., Ziean,L.L. and Kimmel,B.  
 Sequencing of Drosophila chromosome 2L, region 26C1-26D2  
 Unpublished (1997)  
 2 (bases 1 to 209071)  
 Celiker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,  
 Switskas,R.R., Harris,N.L., Agapayani,A., Arcinas,T.T., Baxter,E.,  
 Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,  
 Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
 Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,  
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
 Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,  
 Yee,A., Zhang,R., Ziean,L.L. and Kimmel,B.E.  
 Direct Submission  
 Submitted (15-JUL-1998) Berkeley Drosophila Genome Project, MS

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

COMMENT  
 Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory, MS 64-121  
 Berkeley, CA 94720  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive web site (<http://fruitfly.berkeley.edu/sequence/>) or send  
 email to [drosophila@hgsc.lbl.gov](mailto:drosophila@hgsc.lbl.gov).  
 Library locations: 25-2, 137-57, 12-2.  
 Location/Qualifiers  
 1..209071  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /chromosome="2L"  
 /map="26C1-26D2"  
 /clone="Pls DS00121 (D128), DS05470 (D270), and DS00108  
 (D120)"  
 /note="DS00121 (d128) extends from bp 1 to bp 84231,  
 DS05470 (d270) extends from bp 64,751 to bp 139,027 and  
 DS00108 (d120) sequenced as a bridge, extends from bp  
 137,172 to bp 209,071.

BASE COUNT 59585 a 45615 c 45383 g 58488 t  
 ORIGIN  
 Query Match 1.5%; Score 20; DB 22; Length 209071;  
 Best Local Similarity 100.0%; Pred. No. 3.01e+00;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 179069 CACGATCGCTTACATPAA 179088  
 ||||||||||||||||||  
 Cp 1114 CACGATCGCTTACATPAA 1095

RESULT 13  
 LOCUS AC000537 211037 bp DNA HTG 22-NOV-1998  
 DEFINITION Homo sapiens clone GS308H05, WORKING DRAFT SEQUENCE, 6 unordered  
 pieces.  
 AC000537  
 AC000537  
 93907476  
 AC000537.1 GI:3907476  
 HTG: HTGS\_PHASE1.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 211037)  
 Waterston,R.H.  
 The sequence of Homo sapiens clone  
 Unpublished  
 2 (bases 1 to 211037)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (25-NOV-1998) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 COMMENT  
 On Nov 22, 1998 this sequence version replaced gi:3451378.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1  
 2142: contig of 2142 bp in length  
 2143  
 2160: gap of unknown length  
 16553: contig of 14393 bp in length  
 16554  
 16572: gap of unknown length  
 16572  
 30233: contig of 13662 bp in length  
 30234  
 30251: gap of unknown length  
 59021: contig of 28770 bp in length



FEATURES

\* 59022 59039: gap of unknown length

\* 59040 122511: contig of 63472 bp in length

\* 122512 122529: gap of unknown length

\* 122530 211037: contig of 88508 bp in length.

Location/Qualifiers

SOURCE

1..211037

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="GS308H05"

BASE COUNT 65324 a 42490 c 42250 g 60883 t 90 others

ORIGIN

Query Match 1.5%; Score 20; DB 19; Length 211037;

Best Local Similarity 100.0%; Pred. No. 3.01e+00;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 155301 AATTGAAATTTTGAGCAT 155320

CP 729 AATTGAAATTTTGAGCAT 710

RESULT 14

LOCUS G09531 524 bp DNA STS 14-AUG-1995

DEFINITION human STS CHLC.GCT3B06.P11066 clone GCT3B06.

ACCESSION G09531

NID 9941380

VERSION G09531.1 GI:941380

KEYWORDS STS sequence; primer; sequence tagged site.

SOURCE human vector-pUC1 host-E.coli dut-tung+ (DH10B) Marker Selected

ORGANISM genomic DNA prepared from XY individual of French nationality.

Homo sapiens

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;

Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;

Catarrhini; Homidae; Homo.

1 (bases 1 to 524)

Murray, J., Sheffield, V., Weber, J.L., Duyk, G. and Buettow, K.H.

Cooperative Human Linkage Center

Unpublished (1995)

Synonyms: GCT3B06, CHLC.GCT3B06.P11065

Contact: Dr. Jeffrey C. Murray

UofI

The University of Iowa

Department of Pediatrics, Iowa City, IA 52242, USA

Tel: (319) 356-3508

Fax: (319) 356-3347

Email: jeff-murray@uiowa.edu

Primer A: AATGACTTGAGATGATCG

Primer B: TCCTTAGGCTTGCTTCCT

STS size: 141

PCR Profile:

denature: 30 seconds at 94 degrees C

annealing: 75 seconds at 55 degrees C

extension: 15 seconds at 72 degrees C

PCR cycles: 27

extension: 6 minutes at 72 degrees C

Protocol:

Template: 30ng genomic DNA

Primer: each 1.5 pmole

dNTPs: each 200 uM

Taq Polymerase: 0.3 units

Total Vol: 10 uL

Buffer:

MgCl2: 1.5mM

KCl: 50mM

Tris: 10mM

pH: 8.3.

Location/Qualifiers

1..524

/organism="Homo sapiens"

STS

primer\_bind 90..109

complement(211..230)

BASE COUNT 73 a 101 c 110 g 122 t 118 others

ORIGIN

SOURCE

1.4%; Score 19; DB 34; Length 524;

Best Local Similarity 86.4%; Pred. No. 1.46e+01;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 356 AACAAATTTGATATCTGTGT 1256

QY 1235 AACAAATTTGATATCTGTGT 1256

RESULT 15

LOCUS HUMIGV3A 2650 bp DNA PRI 06-DEC-1994

DEFINITION Human germline Igh chain (hv3005) V3-region.

ACCESSION M83134

NID 9186232

VERSION M83134.1 GI:186232

KEYWORDS V-region; germline; immunoglobulin heavy chain.

SOURCE Homo sapiens (tissue library: Y79-EMBL-3) DNA.

ORGANISM Homo sapiens

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 2650)

Chen, P.P.

Structural analyses of human developmentally regulated Vh3 genes

Scand. J. Immunol. 31 (3), 257-267 (1990)

90208163

Location/Qualifiers

1..2650

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_line="Y79 retinoblastoma"

/tissue\_lib="Y79-EMBL-3"

/map="14q32.33"

/map="14q32.33"

/gene="IGHV"

/note="octamer recombination signal; G00-128-528"

join(1782..1827,1929..1939)

/gene="IGHV"

join(1782..1827,1929..1939)

/note="G00-128-528"

join(1782..1827,1929..2233)

/partial

/gene="IGHV"

/codon\_start=1

/db\_xref="GDB:G00-128-528"

/product="immunoglobulin heavy chain V-region"

/protein\_id="AA56827.1"

/db\_xref="PID:9186233"

/db\_xref="GI:186232"

/translation="MERGLSWFLVALLRVCQVQVVEGGGVQGRRLRSCAS

GFITSSYAMHWROAPGKGLWVAVISYDGSNNRYADSVKGRFTISDNKNILYQM

NSLRAEDTAVYCAR"

1929..2233

/gene="IGHV"

/note="G00-128-528"

1940..2233

/gene="IGHV"

/note="G00-128-528"

/product="immunoglobulin heavy chain V-region"

2236..2242

/gene="IGHV"

/note="heptamer recombination signal; G00-128-528"

2266..2274

/gene="IGHV"

/note="nonamer recombination signal; G00-128-528"

BASE COUNT 769 a 528 c 591 g 762 t

ORIGIN chromosome 14.



Query Match 1.48; Score 19; DB 29; Length 2650;  
 Best Local Similarity 100.08; Pred. No. 1.46e+01;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 1216 TGCATATTAGATAAATA 1234  
 ||||||||||||||||  
 QY 1296 TGCATATTAGATAAATA 1314

Search completed: Sat Nov 27 13:30:14 1999  
 Job time : 2452 secs.



**THIS PAGE BLANK (USPTO)**



\*\*\*\*\*  
 N O T E S  
 \*\*\*\*\*  
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_un n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 14:14:48 1999; Maspar time 8.72 Seconds

Tabular output not generated. 460,971 Million cell updates/sec

Title: >US-09-103-287-1  
 Description: (1-1351) from US09103287.seq  
 Perfect Score: 1351  
 N.A. Sequence: 1 ATGAGTAAGAGCTTTATAT.....TTAATATGTTATATAGAG 1351  
 Comp: TACCATTCCTCAATAATATA.....AATATCAATAATATATCTC

Scoring table: TABLE jmetric  
 Gap 60

Mismatch STD: Dbase 0; Query 0

Searched: 1052 seqs, 1486975 bases x 2

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: HIV-NAB  
 1:H\_PRI 2:H\_DNA 3:H\_VIR

Statistics: Mean 10.393; Variance 2.033; scale 5.112

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	16	1.2	330	3	HIVU04910	Human immunodeficiency 6.32e-01
2	16	1.2	10036	3	STVSABIC	Simian immunodeficiency 6.32e-01
3	15	1.1	1459	3	HIVY169	Human immunodeficiency 4.03e+00
4	15	1.1	9076	2	HIVHAN	Human immunodeficiency 4.03e+00
5	15	1.1	9468	3	FIVPPR	Feline immunodeficiency 4.03e+00
6	15	1.1	9811	3	STVCP2	Simian immunodeficiency 4.03e+00
7	13	1.0	253	3	HIVBOS1	Human immunodeficiency 8.65e+01
8	14	1.0	273	3	HIVU08687	Human immunodeficiency 8.65e+01
9	13	1.0	302	3	HIVFLQ7229	Human immunodeficiency 8.65e+01
10	13	1.0	305	3	HIVFLQ7214	Human immunodeficiency 8.65e+01
11	13	1.0	312	3	HIVFLQ626	Human immunodeficiency 8.65e+01
12	13	1.0	319	3	HIVMOM22	Human immunodeficiency 8.65e+01
13	13	1.0	320	3	HIVFLQ7216	Human immunodeficiency 8.65e+01
14	13	1.0	320	3	HIVFLQ7223	Human immunodeficiency 8.65e+01
15	13	1.0	323	3	HIVFLQ7227	Human immunodeficiency 8.65e+01
16	13	1.0	324	3	HIVFLPE26	Human immunodeficiency 8.65e+01
17	13	1.0	324	3	HIVFLPE22	Human immunodeficiency 8.65e+01
18	13	1.0	324	3	HIVFLPE210	Human immunodeficiency 8.65e+01
19	13	1.0	324	3	HIVFLPE21	Human immunodeficiency 8.65e+01
20	13	1.0	341	3	HIVFLQ624	Human immunodeficiency 8.65e+01

c	21	13	1.0	343	3	HIVFLPD24	Human immunodeficiency 8.65e+01
c	22	14	1.0	345	3	HIVU08774	Human immunodeficiency 2.09e+01
c	23	13	1.0	352	3	HIVFLPD25	Human immunodeficiency 8.65e+01
c	24	13	1.0	367	3	HIVFLPD214	Human immunodeficiency 8.65e+01
c	25	13	1.0	371	3	HIVFLPC212	Human immunodeficiency 8.65e+01
c	26	13	1.0	371	3	HIVFLPD28	Human immunodeficiency 8.65e+01
c	27	13	1.0	675	3	HIVART152	Human immunodeficiency 8.65e+01
c	28	14	1.0	1116	3	SIVVER1E	Simian immunodeficiency 2.09e+01
c	29	13	1.0	1366	3	HIVSB	Human immunodeficiency 8.65e+01
c	30	13	1.0	1462	3	HIVBVS217	Equine immunodeficiency 2.09e+01
c	31	14	1.0	2580	3	ETAVWU5	Equine arthritis ence 2.09e+01
c	32	13	1.0	2589	3	HIV2ALI	Human immunodeficiency 8.65e+01
c	33	13	1.0	2806	3	HIVU08801	Human immunodeficiency 8.65e+01
c	34	13	1.0	3201	3	HIVBAL2	Human immunodeficiency 8.65e+01
c	35	14	1.0	4549	3	HSPUEV	Human spumaretrovirus 2.09e+01
c	36	14	1.0	4549	3	HSPUEV	Human spumaretrovirus 2.09e+01
c	37	14	1.0	7817	2	HIV2D205	Human immunodeficiency 2.09e+01
c	38	14	1.0	8344	3	ETAV	Equine infectious ane 2.09e+01
c	39	13	1.0	8391	3	BIV106	Bovine immunodeficiency 8.65e+01
c	40	13	1.0	9215	2	SIVMNDGB1	Simian immunodeficiency 8.65e+01
c	41	14	1.0	9540	3	HIVJRC5F	Human immunodeficiency 2.09e+01
c	42	14	1.0	9597	3	STVSYK	Simian immunodeficiency 2.09e+01
c	43	14	1.0	9623	3	SIVAGM677	Simian immunodeficiency 2.09e+01
c	44	13	1.0	9672	2	HIV2ST	Human immunodeficiency 8.65e+01
c	45	14	1.0	10271	3	HIV20C1	Human immunodeficiency 2.09e+01

#### ALIGNMENTS

RESULT 1  
 LOCUS HIVU04910 330 bp ds-DNA VRL 01-APR-1994  
 DEFINITION Human immunodeficiency virus type 1, isolate 662 from Baltimore, U.S., partial env cds, C2V3 region.  
 ACCESSION U04910  
 SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate 662, Baltimore, U.S.; cultured on PBMCs.  
 REFERENCE 1 (bases 1 to 330)  
 AUTHORS NIAID/NIH DAIDS Variation Program.  
 COMMENT This sample is part of a set of sequences generated through the NIAID/NIH DAIDS HIV variation program. The virus was derived from an asymptomatic individual, from Baltimore, U.S., whose route of infection is thought to be due to homosexual contact. The blood sample was taken in 1992. This env sequence clusters with HIV-1 B subtype sequences. The full name of this sequence is HIV19ZUS662A1BA.01d1SEB; it was presented in alignments in an abbreviated form in the April 94 Human Retro. AIDS compendium update as B2US662D.01d19ED.  
 Location/Qualifiers  
 CDS  
 <1..>330  
 /note="envelope (AA at 1)"  
 BASE COUNT 142 a 43 c 61 g 84 t  
 ORIGIN 835 bp downstream from beginning of env cds

Query Match 1.2%; Score 16; DB 3; Length 330;  
 Best Local Similarity 100.0%; Pred. No. 6.32e-01;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

db 222 agaacatttgataat 237  
 QY 1233 AGAACAAATTGATAT 1248

#### RESULT 2

LOCUS STVSABIC 10036 bp ss-RNA VRL 25-MAY-1994  
 DEFINITION Simian immunodeficiency virus (SIVsgm) complete genomic sequence, African green monkey isolate SIVsgmAB-1.  
 ACCESSION U04005  
 SOURCE Simian immunodeficiency virus PCR-amplified DNA extracted from cocultured PBMCs derived from a wild female asymptomatic saebus African green monkey captured in its natural habitat of Senegal.  
 ORGANISM Simian immunodeficiency virus  
 Viridae: ss-RNA enveloped viruses; Positive strand RNA virus;







Db 3712 tgatatcaaaatta 3727  
 |||||  
 QY 1275 TGATATCAAAATTA 1290

RESULT 3  
 LOCUS HIV169 1459 bp ss-RNA VRL 20-APR-1993  
 DEFINITION Human immunodeficiency virus type 1, isolate V169 taken from a Rwandan national residing in Belgium, gag region.  
 L11796  
 ACCESSION Human immunodeficiency virus type 1 (HIV-1), Rwandan isolate V169.  
 SOURCE Human immunodeficiency virus type 1  
 ORGANISM Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;  
 Retroviridae; Lentivirinae.  
 REFERENCE Louwaghe, J.J., McCutchan, F., Brennan, T., Peeters, M., Brennan, T., Sanders-Buell, E., Eddy, G., van der Groen, G., Franssen, K., Gershby-Danet, M., Delays, R. and Burke, D.  
 Phylogenetic analysis of gag genes from seventy international HIV-1 isolates provides evidence for multiple genotypes  
 AIDS 7, 769-780 (1993)

JOURNAL  
 STANDARD  
 COMMENT

TITLE Kindly provided prior to publication by Henry M. Jackson Foundation Research Laboratory, 1500 East Gude Drive, Rockville, MD 20850. The authors group the outlying V169 gag sequence with their subtype E. The naming of this subtype was made independently of envelope subtype studies, and so gag sequences in this subtype are not necessarily related to envelope sequences classified as subtype E in the 1992 and 1993 compendiums.

FEATURES  
 CDS  
 1..1459  
 Location/Qualifiers  
 1..1459  
 /product="gag protein"  
 /gene="gag"  
 /codon\_start=1  
 /translation="MGARASVLSGGKLDAMERILRPGGRKKYKKHILIMARELERE  
 ALDRLEETSECRKIRIGLOLSLOSGSELISXIVAVLVFVHQVYKRTKEAE  
 KLEEVNKSQOKQQAADKGYSONPIVONLOGVHVAISPTILNAVKTIEKAF  
 SPEVLPMEFALSAGATPODLNMLNTVGSHQAMOLKDTINEEAEMWDLHPVAGP  
 NPQGMREPRGSDIGTSTLQEOIOWMTSNPIPGDILYKIKIILGLKRIYMSVP  
 SLIDIRGKPEPRDVFVDFPTLLRAEATQVSGMGTDTLLIONNPCKTILKMG  
 PGATLEEMTACOGVGGPGHKARVLAETGSAQANNAIMOKSNFGQRTVCFNCGK  
 EGVAKNCAPRKKGCKMGCRGCHQKCTEQANQALGIMSNKGRPNFLOSREP  
 SAPPESFGFREITPSPKQEDBSPLASLXK"  
 1..1459  
 /organism="Human immunodeficiency virus type 1"  
 /isolate="V169"  
 /cell\_type="lymphocyte"  
 /proviral  
 /sequenced\_mol="DNA"  
 /tissue\_type="blood"

BASE COUNT 531 a 277 c 365 g 286 t  
 ORIGIN gag cds start

Query Match 1.1%; Score 15; DB 3; Length 1459;  
 Best Local Similarity 100.0%; Pred. No. 4.03e+00;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 300 tttagagaagcttga 314  
 |||||  
 QY 849 TTTAGAGAGCTTGA 863

RESULT 4  
 LOCUS HIVAN 9076 bp ss-RNA UNA 20-DEC-1989  
 DEFINITION Human immunodeficiency virus type 1, isolate HAN-2; complete genome.  
 SOURCE HIV-1, isolate HAN-2, clones 2/2 and 2/3; infectious proviral DNA.  
 REFERENCE 1 (bases 1 to 9076)  
 Sautermann, U., Schneider, J., Mous, J., Bruckhorst, U., Schedel, I., Jentsch, K.D. and Hunsmann, G.  
 Molecular cloning and characterization of a German HIV-1 isolate  
 AIDS Res. and Hum. Retroviruses submitted, February 1990

JOURNAL  
 STANDARD

COMMENT Kindly submitted in computer readable form by F. Kirchhoff, DPZ, Göttingen, 2/2/90.

Clone HAN2/3, from which most of the sequence shown was derived, is inactive perhaps as the result of mutations leading to frameshifts in the gp41 portion of the env cds. Clone 2/2 is infectious (in MT-2 cells) and its env sequence has been inserted in this entry; bases 7100-7608 (annotated below).

FEATURES  
 from to/span description  
 pept 259 1758 gag polypeptin  
 pept < 1554 4562 pol polypeptin  
 pept 4507 5085 vif protein  
 pept 5025 5315 vpr protein  
 pept 5296 5510 tat, exon 2 (first expressed exon)  
 pept 7841 7931 tat, exon 3 (AA at 7842)  
 pept 5435 5510 rev, exon 2 (first expressed exon)  
 7841 8115 rev, exon 3 (AA at 7843)  
 5527 5772 vpr protein  
 pept 5690 8257 env polypeptin  
 variation 7283 8257 cg in 2/2; c in 2/3  
 variation 7293 3 cgc in 2/2; c in 2/3  
 variation 7325 2 g9 in 2/2; g in 2/3  
 variation 7364 1 g in 2/2; a in 2/3  
 variation 7494 1 t in 2/2; c in 2/3  
 pept 8259 8861 nef protein  
 BASE COUNT 3276 a 1595 c 2186 g 2019 t  
 ORIGIN 5' terminus of LTR US region

Query Match 1.1%; Score 15; DB 2; Length 9076;  
 Best Local Similarity 100.0%; Pred. No. 4.03e+00;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5210 aatactgcaaacact 5224  
 |||||  
 Cp 1040 AATACGTCAACAACACT 1026

RESULT 5  
 LOCUS FIVPR 9468 bp ss-RNA VRL 23-JUL-1990  
 DEFINITION Feline immunodeficiency virus, PPR isolate, complete genome; (San Diego isolate).  
 ACCESSION M36968  
 SOURCE Feline immunodeficiency virus, clone PPR, genomic DNA. Infectious clone.  
 REFERENCE 1 (bases 1 to 9468)  
 Phillips, T.R., Talbot, R.L., Lamont, C., Muir, S., Lovelace, K. and Elder, J.H.  
 Comparison of two host cell range variants of feline immunodeficiency virus  
 J. Virol. 64, 4605-4613 (1990)

JOURNAL  
 STANDARD  
 COMMENT

Kindly provided in computer readable form by T. Phillips, Scripps Research Foundation, La Jolla CA.

The PPR isolate is approximately 91% homologous to the previously characterized Petaluma isolate. The latter infected feline kidney cells and the G355-5 cell line but replicated less efficiently on feline PBL's. In contrast, PPR productively infects PBL's but not the other cell lines. The authors point out interesting differences in the LTRs and coding regions.

The small orfs include: orf 1, in size and position similar to vif like; orfs 3 (denoted D in [1]) and 4 (denoted H in [1]) seen also in the Petaluma strain.  
 Location/Qualifiers  
 628..1980  
 /note="gag polypeptin"  
 /codon\_start=628  
 <1869..5243  
 /partial  
 /note="pol polypeptin (NH2-terminus uncertain; AA at



```

RESULT      7          HIVB051           253 bp ss-RNA          VRL       24-AUG-1989
LOCUS              Human immunodeficiency virus type 1, isolate 1153, envelope region
DEFINITION      M2910
ACCESSION
SOURCE          Human immunodeficiency virus type 1 (HIV-1), isolate 1153, clone
REFERENCE      1 (bases 1 to 253)
AUTHORS        Ou,C.-Y.
TITLE          ;
JOURNAL        unpublished (1989)
STANDARD
COMMENT
FEATURES
BASE COUNT      Approximately 1170 nt from env start (BRU coordinates)
ORIGIN
Query Match            1.0%: Score 13; DB 3; Length 253;
Best Local Similarity 100.0%; Pred.No. 8.65e+01;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db    160 tagatgttcacaa 172
        |||||||||
CP    648 TAGATGTCATCA 636

RESULT      8          HIVU0687           273 bp ss-RNA          VRL       01-APR-1994
LOCUS              Human immunodeficiency virus type 1, sample 017 from Brazil,
DEFINITION      C2V3 of .env cds.
ACCESSION      U0687
SOURCE          Human immunodeficiency virus type 1 (HIV-1), sample 017 from
                Brazil; derived from a primary isolate.
REFERENCE      1 (bases 1 to 273)
AUTHORS        WHO Global Programme on AIDS.
COMMENT        This sample is part of a set of sequences generated through the WHO
                Global Programme on AIDS. The virus was derived from an
                asymptomatic individual, from Brazil, whose route of infection is
                thought to be due to homosexual contact. The blood sample was
                taken in 1992. This env sequence clusters with HIV-1 B subtype
                sequences. The full name of this sequence is
                HVI92BR01WHO_01__1gCR; it was presented in alignments in an
                abbreviated form in the April 94 Human Retroviruses and AIDS
                compendium update as B2BR017W_01__1gCR.
                location/Qualifiers
                <1..>273
                /note='env polyprotein (AA at 1)'
BASE COUNT      129 a         41 c         45 g         58 t
ORIGIN          811 bp downstream from the beginning of env cds
Query Match            1.0%: Score 14; DB 3; Length 273;
Best Local Similarity 100.0%; Pred.No. 2.09e+01;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db    242 aacaattgtcaaat 255
        |||||||||
OY    1235 AACAAATTGTGAATA 1248

RESULT      9          HIVFLD7229          302 bp ss-RNA          VRL       03-FEB-1992
LOCUS              Human immunodeficiency virus type 1, viral sample LC03B29, V4C3V5
DEFINITION      Human immunodeficiency virus type 1,
                region.
ACCESSION      M9149
SOURCE          Human immunodeficiency virus type 1 (HIV-1), M13 clone B29 of
                a Florida local control (LC03); DNA ID 5244.
ORGANISM        Human immunodeficiency virus 1

```



REFERENCE  
AUTHORS

Viridae: ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
1 (bases 1 to 302)  
Ou,C.-Y., Ciesielski,C.A., Myers,G., Bandea,C.I., Luo,C.-C.,  
Korber,B.T.M., Mullins,J.I., Schochetman,G., Berkelman,R.L.,  
Economou,A.N., Witte,J.D., Furman,L.J., Satten,G.A., MacInnes,K.A.,  
Curran,J.W., Jaffe,H.W., et al.  
Molecular Epidemiology of HIV Transmission in a Dental Practice  
Science (1992) In press

TITLE  
JOURNAL  
STANDARD  
COMMENT

full staff review  
Kindly submitted in computer readable form by the CDC (Centers for  
Disease Control), Atlanta, GA.

The sequence in this entry is one of 12 clones over the VAC3V5  
region obtained by the CDC from this Florida control sample.  
An alignment presenting all 12 clone sequences from the VAC3V5  
region from this subject follows entry <HIVLQ7227> in the March  
update to the 1991 Human Retroviruses and AIDS.  
The sequences for the VAC3V5 region for this local control were  
generated by running the program "comp" and are the opposite sense  
strand of the sequences originally forwarded to the HIV Sequence  
Database by the CDC.

Please note that for this set of sequences, clone numbers from the  
V3 region do not correspond with similar numbers from the VAC3V5  
region.  
VAC3V5-region sequences include sequences under the following  
accession numbers: M91084 - M91089 (dentist); M91090 - M91115 and  
M91121 - M91131 (dentist's patients); and M91132 - M91156 (Florida  
local controls). V3-region sequences used in [1] include: M90847  
- M90853 (dentist); M90854 - M90912 (dentist's patients); and  
M90914 - M90966 (Florida local controls).  
For related sequences, see also M90913 and M92100 - M92150.

FEATURES  
CDS

<1..>302  
/note="env polypeptide (AA at 3)"  
/codon\_start=3  
/translation="IYMSFNGGFEFFYCNLTQFNSTWTKTGEGSNHTKGDATITLPC  
RIKQIINLMQEVKAMVAPPIRGQIRCSNITGLLTRDGGTNGTETFRPGGN"  
BASE COUNT 113 a 53 c 63 g 73 t  
ORIGIN

Query Match 1.0%; Score 13; DB 3; Length 302;  
Best Local Similarity 100.0%; Pred. No. 8.65e+01;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 209 tagatgttcata 221  
|||||  
CP 648 TAGATGTTTCATCA 636

## RESULT 10

LOCUS HIVLQ7214 305 bp ss-RNA VRL 03-FEB-1992  
DEFINITION Human immunodeficiency virus type 1, viral sample LC03B14, VAC3V5  
region.  
ACCESSION M91145  
SOURCE Human immunodeficiency virus type 1 (HIV-1), M13 clone B14 of  
a Florida local control (LC03), DNA ID 5244.

## ORGANISM

Human immunodeficiency virus type 1 (HIV-1), M13 clone B14 of  
a Florida local control (LC03), DNA ID 5244.  
Viridae: ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
1 (bases 1 to 305)

REFERENCE  
AUTHORS

Korber,B.T.M., Mullins,J.I., Schochetman,G., Berkelman,R.L.,  
Economou,A.N., Witte,J.D., Furman,L.J., Satten,G.A., MacInnes,K.A.,  
Curran,J.W., Jaffe,H.W., et al.  
Molecular Epidemiology of HIV Transmission in a Dental Practice  
Science (1992) In press  
full staff review  
Kindly submitted in computer readable form by the CDC (Centers for  
Disease Control), Atlanta, GA.

The sequence in this entry is one of 12 clones over the VAC3V5

region obtained by the CDC from this Florida control sample.  
An alignment presenting all 12 clone sequences from the VAC3V5  
region from this subject follows entry <HIVLQ7227> in the March  
update to the 1991 Human Retroviruses and AIDS.  
The sequences for the VAC3V5 region for this local control were  
generated by running the program "comp" and are the opposite sense  
strand of the sequences originally forwarded to the HIV Sequence  
Database by the CDC.

Please note that for this set of sequences, clone numbers from the  
V3 region do not correspond with similar numbers from the VAC3V5  
region.  
VAC3V5-region sequences include sequences under the following  
accession numbers: M91084 - M91089 (dentist); M91090 - M91115 and  
M91121 - M91131 (dentist's patients); and M91132 - M91156 (Florida  
local controls). V3-region sequences used in [1] include: M90847  
- M90853 (dentist); M90854 - M90912 (dentist's patients); and  
M90914 - M90966 (Florida local controls).  
For related sequences, see also M90913 and M92100 - M92150.

FEATURES  
CDS

<1..>305  
/note="env polypeptide (AA at 3)"  
/codon\_start=3  
/translation="IYKHSFNGGFEFFYCNLTQFNSTWTKTGEGSNHTKGDATITLPC  
RIKQIINLMQEVKAMVAPPIRGQIRCSNITGLLTRDGGTNGTETFRPGGN"  
BASE COUNT 117 a 57 c 63 g 68 t  
ORIGIN

Query Match 1.0%; Score 13; DB 3; Length 305;  
Best Local Similarity 100.0%; Pred. No. 8.65e+01;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 209 tagatgttcata 221  
|||||  
CP 648 TAGATGTTTCATCA 636

## RESULT 11

LOCUS HIVLQ626 312 bp ss-RNA VRL 03-FEB-1992  
DEFINITION Human immunodeficiency virus type 1, viral sample LC02B6, VAC3V5  
region.  
ACCESSION M91135  
SOURCE Human immunodeficiency virus type 1 (HIV-1), M13 clone B6 of  
a Florida local control (LC02), DNA ID 5242.

## ORGANISM

Human immunodeficiency virus type 1  
Viridae: ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
1 (bases 1 to 312)

Korber,B.T.M., Mullins,J.I., Schochetman,G., Berkelman,R.L.,  
Economou,A.N., Witte,J.D., Furman,L.J., Satten,G.A., MacInnes,K.A.,  
Curran,J.W., Jaffe,H.W., et al.  
Molecular Epidemiology of HIV Transmission in a Dental Practice  
Science (1992) In press  
full staff review  
Kindly submitted in computer readable form by the CDC (Centers for  
Disease Control), Atlanta, GA.

The sequence in this entry is one of 6 clone sequences over the  
VAC3V5 region obtained by the CDC from this Florida control sample.  
An alignment presenting all 6 clone sequences from the VAC3V5  
region from this subject follows entry <HIVLQ6225> in the March  
update to the 1991 Human Retroviruses and AIDS.

All the sequences for the VAC3V5 region for this local control were  
generated by running the program "comp" and are the opposite sense  
strand of the sequences originally forwarded to the HIV Sequence  
Database by the CDC.  
Please note that for this set of sequences, clone numbers from the  
V3 region do not correspond with similar numbers from the VAC3V5  
region.  
VAC3V5-region sequences include sequences under the following  
accession numbers: M91084 - M91089 (dentist); M91090 - M91115 and



M91121 - M91131 (dentist's patients); and M91132 - M91156 (Florida local controls). V3-region sequences used in [1] include: M90847 - M90853 (dentist); M90854 - M90912 (dentist's patients); and M90914 - M90966 (Florida local controls).  
For related sequences, see also M90913 and M92100 - M92150.

## FEATURES

Location/Qualifiers  
<1..>312  
/note="env polyprotein (AA at 1)"  
/codon\_start=1  
/translation="GDPPIVHSENGGEFFRYNTTOLFSTWNGTSGNSSTKNDITLPCKRIQIINMGEVKAMYPPIEGQIRCSNITGILLTRDGGNNNSTMETFPFGCGN"

## BASE COUNT

118 a 55 c 69 g 70 t

Query Match 1.0%; Score 13; DB 3; Length 312;  
Best Local Similarity 100.0%; Pred. No. 8.65e+01;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 219 tagatgttcacatca 231  
|||||  
Cp 648 TAGATGTTTCATCA 636

RESULT 12  
LOCUS HIWOM22 319 bp ss-RNA VRL 14-FEB-1992  
DEFINITION Human immunodeficiency virus type 1, partial env cds, V4V5 region, clone CM560-491R.  
ACCESSION M76925 M76913 M76914 M76915 M76916 M76917 M76918 M76919 M76920 M76921 M76922 M76923 M76924  
SOURCE Human immunodeficiency virus type 1 (HIV-1) PCR amplified proviral DNA from peripheral blood, clone CM560-491R.  
SEGMENT 2 of 2  
REFERENCE 1 (bases 1 to 319)  
AUTHORS Molinsky,S.M., Witte,C.M., Korber,B.T.M., Hutto,C., Parks,W.P., Rosenblum,L.L., Kunstman,K.J., Furtado,M.R. and Munoz,J.L.  
TITLE Selective transmission of human immunodeficiency virus type 1 variants from mothers to infants  
JOURNAL Science 255, 1134-1136 (1992)

## JOURNAL STANDARD

The sequence in this entry is one of 13 sequences over the V4V5 region obtained from this patient. This sequence was selected as representative of the set on the basis of its closeness to the consensus.

An alignment of all 13 sequences with the sequences from the paired infant (HIVOT22) follows entry <HIVOT22> in the March update of Human Retroviruses and AIDS, 1992.

FEATURES  
pept < 1 > 319 envelope protein, V4V5 region (partial; AA at 2)  
BASE COUNT 111 a 47 c 77 g 84 t  
ORIGIN 60 bp upstream of V4 region cds

Query Match 1.0%; Score 13; DB 3; Length 319;  
Best Local Similarity 100.0%; Pred. No. 8.65e+01;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 220 tagatgttcacatca 232  
|||||  
Cp 648 TAGATGTTTCATCA 636

RESULT 13  
LOCUS HIVEL07216 320 bp ss-RNA VRL 03-FEB-1992  
DEFINITION Human immunodeficiency virus type 1, viral sample LC03B16, V4C3V5 region.  
ACCESSION M91146  
SOURCE Human immunodeficiency virus type 1 (HIV-1), M13 clone B16 of a Florida local control (LC03), DNA ID 5244.  
ORGANISM Human immunodeficiency virus type 1  
Virus: ss-RNA enveloped viruses; Positive strand RNA virus; Retroviridae; Lentivirinae.

REFERENCE 1 (bases 1 to 320)  
AUTHORS Ou,C.-Y., Ciesielski,C.A., Myers,G., Bandea,C.I., Luo,C.-C.,

Korber,B.T.M., Mullins,J.I., Schochetman,G., Berkelman,R.L., Economou,A.N., Witte,J.J., Furman,L.J., Satten,G.A., MacInnes,K.A., Curran,J.W., Jaffe,H.W., et al.  
Molecular Epidemiology of HIV Transmission in a Dental Practice Science (1992) In press  
full staff\_review  
STANDARD  
TITLE  
JOURNAL  
COMMENT

Kindly submitted in computer readable form by the CDC (Centers for Disease Control), Atlanta, GA.

The sequence in this entry is one of 12 clones over the V4C3V5 region obtained by the CDC from this Florida control sample. An alignment presenting all 12 clone sequences from the V4C3V5 region from this subject follows entry <HIVEL07227> in the March update to the 1991 Human Retroviruses and AIDS.

The sequences for the V4C3V5 region for this local control were generated by running the program "comp" and are the opposite sense strand of the sequences originally forwarded to the HIV Sequence Database by the CDC.

Please note that for this set of sequences, clone numbers from the V3 region do not correspond with similar numbers from the V4C3V5 region.

V4C3V5-region sequences include sequences under the following accession numbers: M91084 - M91089 (dentist); M91090 - M91115 and M91121 - M91131 (dentist's patients); and M91132 - M91156 (Florida local controls). V3-region sequences used in [1] include: M90847 - M90853 (dentist); M90854 - M90912 (dentist's patients); and M90914 - M90966 (Florida local controls).  
For related sequences, see also M90913 and M92100 - M92150.

## FEATURES

Location/Qualifiers  
<1..>320  
/note="env polyprotein (AA at 3)"  
/codon\_start=3  
/translation="VHSENGGEFFRYNTTOLFSTWNGTSGNSSTKNDITLPCKRIQIINMGEVKAMYPPIEGQIRCSNITGILLTRDGGNNNSTMETFPFGCGGD"

BASE COUNT 114 a 59 c 71 g 76 t

Query Match 1.0%; Score 13; DB 3; Length 320;  
Best Local Similarity 100.0%; Pred. No. 8.65e+01;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 224 tagatgttcacatca 236  
|||||  
Cp 648 TAGATGTTTCATCA 636

RESULT 14  
LOCUS HIVEL07223 320 bp ss-RNA VRL 03-FEB-1992  
DEFINITION Human immunodeficiency virus type 1, viral sample LC03B23, V4C3V5 region.  
ACCESSION M91147  
SOURCE Human immunodeficiency virus type 1 (HIV-1), M13 clone B23 of a Florida local control (LC03), DNA ID 5244.  
ORGANISM Human immunodeficiency virus type 1  
Virus: ss-RNA enveloped viruses; Positive strand RNA virus; Retroviridae; Lentivirinae.

REFERENCE 1 (bases 1 to 320)  
AUTHORS Ou,C.-Y., Ciesielski,C.A., Myers,G., Bandea,C.I., Luo,C.-C., Korber,B.T.M., Mullins,J.I., Schochetman,G., Berkelman,R.L., Economou,A.N., Witte,J.J., Furman,L.J., Satten,G.A., MacInnes,K.A., Curran,J.W., Jaffe,H.W., et al.  
Molecular Epidemiology of HIV Transmission in a Dental Practice Science (1992) In press  
full staff\_review  
STANDARD  
TITLE  
JOURNAL  
COMMENT

Kindly submitted in computer readable form by the CDC (Centers for Disease Control), Atlanta, GA.  
The sequence in this entry is one of 12 clones over the V4C3V5 region obtained by the CDC from this Florida control sample. An alignment presenting all 12 clone sequences from the V4C3V5 region from this subject follows entry <HIVEL07227> in the March



update to the 1991 Human Retroviruses and AIDS.  
The sequences for the V4C3V5 region for this local control were generated by running the program "comp" and are the opposite sense strand of the sequences originally forwarded to the HIV Sequence Database by the CDC.

Please note that for this set of sequences, clone numbers from the V3 region do not correspond with similar numbers from the V4C3V5 region.

V4C3V5-region sequences include sequences under the following accession numbers: M91084 - M91089 (dentist); M91090 - M91115 and M91121 - M91131 (dentist's patients); and M91132 - M91156 (Florida local controls). V3-region sequences used in [1] include: M90847 - M90853 (dentist); M90854 - M90912 (dentist's patients); and M90914 - M90966 (Florida local controls).  
For related sequences, see also M90913 and M92100 - M92150.

## FEATURES

CDS

<1..>320  
/note="env polypeptide (AA at 3)"  
/codon\_start=3  
/translation="IVMHSFNCGGEFFYCNTOLFNSTWNTDRTNNIVNDYGNNDT  
LILPCRIKQIIMMOMQEVGKAMYPPIRGQIRCSSNITGLLTRDGGTNGTETTER  
PGGSD"

BASE COUNT 124 a 55 c 66 g 75 t  
ORIGIN

## Query Match

Best Local Similarity 100.0%; Pred. No. 8.65e+01; Length 320;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 224 tagatgtcatca 236  
|||||

Cp 648 TAGATGTCATCA 636

## RESULT 15

LOCUS HIVFLQ7227 323 bp ss-RNA VRL 03-FEB-1992  
DEFINITION Human immunodeficiency virus type 1, viral sample IC03B27, V4C3V5 region.  
M91148

ACCESSION Human immunodeficiency virus type 1 (HIV-1), M13 clone B27 of a Florida local control (IC03), DNA ID 5244.

SOURCE Human immunodeficiency virus type 1  
Virus; ss-RNA enveloped viruses; Positive strand RNA virus;

ORGANISM Retroviridae; Lentivirinae.  
1 (bases 1 to 323)

REFERENCE Ou,C.-Y., Ciesielski,C.A., Myers,G., Bandea,C.I., Luo,C.-C.,  
Korber,B.T.M., Mullins,J.L., Schochetman,G., Berkeiman,R.L.,  
Economou,A.N., Witte,J.J., Furman,L.J., Satten,G.A., MacInnes,K.A.,  
Curran,J.W., Jaffe,H.W., et al.

TITLE Molecular Epidemiology of HIV Transmission in a Dental Practice  
JOURNAL Science (1992) In press

STANDARD full staff-review

COMMENT kindly submitted in computer readable form by the CDC (Centers for Disease Control), Atlanta, GA.

The sequence in this entry is one of 12 clones over the V4C3V5 region obtained by the CDC from this Florida control sample. An alignment presenting all 12 clone sequences from the V4C3V5 region from this subject follows entry <HIVFLQ7227> in the March update to the 1991 Human Retroviruses and AIDS.  
The sequences for the V4C3V5 region for this local control were generated by running the program "comp" and are the opposite sense strand of the sequences originally forwarded to the HIV Sequence Database by the CDC.

Please note that for this set of sequences, clone numbers from the V3 region do not correspond with similar numbers from the V4C3V5 region.  
V4C3V5-region sequences include sequences under the following accession numbers: M91084 - M91089 (dentist); M91090 - M91115 and M91121 - M91131 (dentist's patients); and M91132 - M91156 (Florida local controls). V3-region sequences used in [1] include: M90847

- M90853 (dentist); M90854 - M90912 (dentist's patients); and M90914 - M90966 (Florida local controls).  
For related sequences, see also M90913 and M92100 - M92150.

## FEATURES

CDS

<1..>323  
/note="env polypeptide (AA at 3)"  
/codon\_start=3  
/translation="IVMHSFNCGGEFFYCNTOLFNSTWNTDRTNNIVNDYGNNDT  
LILPCRIKQIIMMOMQEVGKAMYPPIRGQIRCSSNITGLLTRDGGTNGTETTER  
PGGSD"

BASE COUNT 120 a 58 c 68 g 77 t  
ORIGIN

## Query Match

Best Local Similarity 100.0%; Pred. No. 8.65e+01; Length 323;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 227 tagatgtcatca 239  
|||||

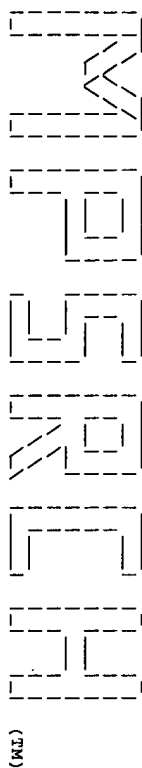
Cp 648 TAGATGTCATCA 636

Search completed: Sat Nov 27 14:15:00 1999  
Job time : 12 secs.



**THIS PAGE BLANK (USPTO)**





Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*

Run on: Sat Nov 27 14:09:15 1999; MasPar time 306.12 Seconds

Tabular output not generated. 945.649 Million cell updates/sec

Title: >US-09-103-287-1  
Description: (1-1351) from US09103287.seq  
Perfect Score: 1351  
N.A. Sequence: 1 ATGAGTACGAGCTTTATAT.....TTAATATGTTTATATATGAG 1351  
Comp: TACTCATTCCTCAAAATATA.....AATTATACAAATATATCTC

Scoring table: TABLE jmetric  
Gap 60

Match STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: n-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39 40:part40 41:part41 42:part42 43:part43  
44:part44 45:part45 46:part46 47:part47 48:part48  
49:part49 50:part50 51:part51 52:part52 53:part53  
54:part54 55:part55 56:part56 57:part57 58:part58  
59:part59 60:part60

Statistics: Mean 7.732; Variance 3.352; scale 2.307

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1351	100.0	1351	60	V99650	UDP-N-acetylmuramate:	0.00e+00
2	921	68.2	2424	58	V74703	Staphylococcus aureus	0.00e+00
3	660	48.9	V80065		Partial nucleotide se	0.00e+00	
4	550	40.7	V53479		DNA encoding a Staphy	0.00e+00	
5	32	2.4	4956	60	X13228	Enterococcus faecalis	7.78e-10
6	20	1.5	2883	60	X13223	Enterococcus faecalis	3.60e-01
7	19	1.4	19	60	V99652	Murc polynucleotides	1.57e+00
8	19	1.4	19	60	V99651	Murc polynucleotides	1.57e+00
9	19	1.4	1267	46	V27381	Streptococcus pneumoniae	1.57e+00

C	10	19	1.4	1825	47	V43027	Streptococcus pneumoniae	1.57e+00
C	11	19	1.4	11864	47	V52194	Streptococcus pneumoniae	1.57e+00
C	12	17	1.3	67	6	Q36987	VIL1, Ox VL 5', end PCR	2.60e+01
C	13	18	1.3	154	58	V75600	Staphylococcus aureus	6.54e+00
C	14	17	1.3	458	53	V89109	EST clone B2187.	2.60e+01
C	15	17	1.3	474	8	O59532	Human brain Expressed	2.60e+01
C	16	17	1.3	600	48	V47570	Leishmania antigen Lm	2.60e+01
C	17	17	1.3	787	58	V74757	Staphylococcus aureus	2.60e+01
C	18	17	1.3	853	41	V17143	Human proteolipid (PL	2.60e+01
C	19	17	1.3	870	6	Q36982	Ox VH-hinge-VL insert	2.60e+01
C	20	18	1.3	1102	60	X13709	Enterococcus faecalis	6.54e+00
C	21	17	1.3	1370	2	N80591	Sequence of P-2 gene	2.60e+01
C	22	17	1.3	1491	12	O6726	Comamonas testosteron	2.60e+01
C	23	17	1.3	1620	48	T98590	DNA encoding argB and	2.60e+01
C	24	17	1.3	1791	20	T04403	Heat shock protein co	2.60e+01
C	25	17	1.3	2269	12	O71025	CytII gene which enc	2.60e+01
C	26	17	1.3	2465	24	T38500	Human HP-8 antigen cd	2.60e+01
C	27	18	1.3	2488	58	V74418	Staphylococcus aureus	6.54e+00
C	28	17	1.3	2546	2	N60695	Fragment pHS 53 homol	2.60e+01
C	29	17	1.3	2800	60	V82058	Streptococcus pneumoniae	2.60e+01
C	30	17	1.3	2800	60	V82059	Streptococcus pneumoniae	2.60e+01
C	31	17	1.3	2800	60	V82066	Streptococcus pneumoniae	2.60e+01
C	32	17	1.3	2800	60	V82065	Streptococcus pneumoniae	2.60e+01
C	33	17	1.3	3122	58	V74830	Staphylococcus aureus	2.60e+01
C	34	17	1.3	3207	47	V52323	Streptococcus pneumoniae	2.60e+01
C	35	17	1.3	3359	52	V59078	Streptococcus pneumoniae	2.60e+01
C	36	17	1.3	3628	38	T92868	Soybean seed coat per	2.60e+01
C	37	17	1.3	3628	38	T92701	Candida CARM2 gene.	2.60e+01
C	38	17	1.3	4410	15	O93913	Yeast MS1 gene.	2.60e+01
C	39	17	1.3	4700	47	V30461	Soybean seed coat per	2.60e+01
C	40	17	1.3	5866	7	Q38640	Sequence of exons II	2.60e+01
C	41	17	1.3	6211	47	V52141	Streptococcus pneumoniae	2.60e+01
C	42	17	1.3	16484	60	X13095	Enterococcus faecalis	2.60e+01
C	43	18	1.3	58407	48	V21210	Methanococcus jannasch	6.54e+00
C	44	16	1.2	3905	60	X13036	Enterococcus faecalis	9.82e+01

# ALIGNMENTS

RESULT 1  
ID V99650 standard; DNA; 1351 BP.  
AC V99650.  
DT 17-MAR-1999 (first entry)  
DE UDP-N-acetylmuramate:L-alanine ligase. (Murc polypeptide) encoding DNA.  
KW Murc gene: UDP-N-acetylmuramate:L-alanine ligase; Murc polypeptide;  
KW bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
KW immunogen; drug; genetic immunisation; ds.  
OS Staphylococcus aureus.  
FH Key Location/Qualifiers  
FT CDS 22..1335  
FT /tag= "a"  
FT /gene= "Murc"  
FT /product= "UDP-N-acetylmuramate:L-alanine ligase  
(Murc polypeptide)"  
EP-889123-A2.  
PD 07-JAN-1999.  
PF 26-JUN-1998: 305064.  
PR 03-JUL-1997: US-052720.  
PA (SMK ) SMITHKLINE BEECHAM CORP.  
PA (SMK ) SMITHKLINE BEECHAM PLC.  
PI Burnham MKR, Wallis NG;  
DR WPI: 99-062655/06.  
DR P-PSDB: W87771.  
PT New isolated Murc polypeptide from Staphylococcus aureus and related  
PT nucleic acid - useful in diagnosis, treatment and prevention of  
PT bacterial infections  
PS Clam 2: Pages 3-4; 39p; English.  
CC The present sequence represents a Murc gene encoding a Staphylococcus  
CC aureus UDP-N-acetylmuramate:L-alanine ligase (Murc polypeptide). Host  
CC cell containing an expression system comprising the Murc gene can be used  
CC for the recombinant production of the polypeptide. Agonists or the Murc  
CC polypeptide are used to treat conditions requiring increased activity or  
CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or



CC competitive polypeptide are useful for inhibiting the polypeptide e.g.  
 CC bacterial (especially S. aureus) infections. They are also useful against  
 CC Helicobacter pylori infections and related cancers, ulcers and gastritis.  
 CC The antibacterial agents are useful to treat in-dwelling devices for  
 CC infection prevention or generally as wound treatments to prevent adhesion  
 CC of bacteria to matrix proteins. The Murc polypeptide is also useful for  
 CC diagnosing or prognosing a (susceptibility to) disease, for raising  
 CC antibodies to identify immunogens or specific receptors. In rational  
 CC drug design and as an immunogen for vaccines. The Murc gene sequences are  
 CC useful in antisense/ribozyme therapeutics; to detect mutant Murc gene;  
 CC for chromosomal mapping; to determine bacterial serotype; and for genetic  
 CC immunisation.  
 SQ Sequence 1351 BP; 471 A; 166 C; 260 G; 454 T;  
 Query Match 100.0%; Score 1351; DB 60; Length 1351;  
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
 Matches 1351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 1 atgagtaagggtttatataatagacacacttcttctgtcgaaataaagttctggc 60  
 Qy 1 ATGAGTAAGGAGTTTATATATATGACACATATCTTTGCGAATTAAGGTTCTGGC 60  
 Db 61 atgaattcattagcaacaatcatgcatgtaagacatgaagttcaagatcgatatt 120  
 Qy 61 ATGAGTTTATATGCAACAATCATGTCATGATTTAGGACATGAAGTTCAAGATCGGATATT 120  
 Db 121 gagaactacgtatattacagaagtgctcttagaataaaggagataaataattacatt 180  
 Qy 121 GAGAACTACGTTATACGAAGTGTCTTAGAATAAAGGAGATAAATAATTACATT 180  
 Db 181 ggtgtaataacataaagaagaatgtaagtaatacaagaatgaagcttgcgagtgac 240  
 Qy 181 GGTGTAATAACATAAAGAAGATGTAAGTATACGAAGTAAGCATTCGCGAGTGAC 240  
 Db 241 catgaagaataatgtagcgtgcacatcaattgaattagattgttgaattatatttt 300  
 Qy 241 CATGAAGAATAATGTCGTCACATCAATTGAATTGAGTTGTAATTATATGATTTT 300  
 Db 301 ttaggacagatattatgacataatacttcagtagctgtaactgtgcacatgtaaaact 360  
 Qy 301 TTAGGACAGATATTATGATCAATATATCTTCAGTAGCTGTAACTGTCACATGTAACACT 360  
 Db 361 tctaaacaagtttatattacatgattatgaatggtgataaagaacttatttattt 420  
 Qy 361 TCTAAACAAGGTTTATTTACATGTTTGAATGCGTGAATAAACAATCTTATTTAATT 420  
 Db 421 ggtgtagcacaggtatggtgcttgaagtgattttcgctttgaagcattgaa 480  
 Qy 421 GGTGTCGACAGATATGAGATGCTGTAAGTGAATTTGCTTTGAGGCAATGTGA 480  
 Db 481 tatagaagtcacattttaaagtataaactgtaacgaattatgacaaatattgatttc 540  
 Qy 481 TATAGAGTCACACTTTTAAAGTATTAACCTGATTCGCAATATATGCAAAATGATTTTC 540  
 Db 541 gatcatcctgattatttcaagaatatgaatggtttttgaagcattcccaaaatgaca 600  
 Qy 541 GATCATCCTGATTATTTCAAAATATATGATGATGTTTGAAGCATTCCTCAAAATGACA 600  
 Db 601 cataatgtaaaaaaaggtattatgcttgagggtgtagatgaacatctgaaattgaa 660  
 Qy 601 CATATATGTTAAAAAGGATTTTCTTGCGGTGATGATGAACATCTACGTAAATTTGAA 660  
 Db 661 gcaagatgctcaatttactatagatttaagaattcgatgacattatgctcaaat 720  
 Qy 661 GCAAGATGTTCAATTTATATGATTAAGATTAAGATTCGATGACATTTATCTCAAAAT 720  
 Db 721 attcaaatgaagataaaggtacgtctttagtgtagtgtagtgtagtgtagtgtagt 780  
 Qy 721 ATTCAAAATGAGATTAAGGATCTCTTTGATGTGATGTCGATGTCGATGATTTATGAT 780  
 Db 781 caattccgtctccacaataatgtagacataaagtttaaatgacttgcttaattcg 840  
 Qy 781 CACTTCTCTCTCCACATATGATGATGATGATGATTTTAAATGATTTAGCTGATTTGGC 840

Db 841 attagatttagagaagctagatgattacaatatataaagaagctagaagcttggt 900  
 Qy 841 ATTAGATTTTTAGAGAACCTGATGATTTCAAAATATTAAGAAGCATTTGAAGCTTTGCT 900  
 Db 901 ggtgttaacgctgttttaagaactaacattgcaaatcaattatgtatgatattat 960  
 Qy 901 GGTGTTAAACGCTGTTTAAAGAACTCAATGCAAAATGCAAAATGATTTGATGATTTAT 960  
 Db 961 gcaacacatccaaagaaattagtgctcaaatgacacagacagcaagaagaatcacat 1020  
 Qy 961 GCACACATCCAAAGAAATTAGTGCTACATTTGACACAGCAGCAAAATATCCACAT 1020  
 Db 1021 aaagaagttgttcagatcttcaacacacacacacacacacacacacacacacac 1080  
 Qy 1021 AAAGAAGTTGTTCAGATCTTTCAACACACACACTTCTCTAAGACACAGCATTTTAAAT 1080  
 Db 1081 gaattgcaagaagttatgtaagaagcagatcgtgattcttaigtgaattttggctca 1140  
 Qy 1081 GAATTGCAAGAAGTTATGTAAGCAGATCGTATCTTATGTAAGATTTTGCTCA 1140  
 Db 1141 attagaagaattctggtgcattacagatacaagaattatgataaattggaagtga 1200  
 Qy 1141 ATTAGAGAAATTCGCGGCATTTAAGATTAAGATTAATTAATTAATTAATTAATTAAT 1200  
 Db 1201 tgcgttcaattgaagatcttattatgattatgatacaaatgataatgctgtgtttta 1260  
 Qy 1201 TCGTTCAATTAAGAGATCTTATTAATGATTAAGCAATTTGATTAATGCTGTTTATTA 1260  
 Db 1261 ttatagggtcaggtgattacaaatattacaagaattatgataaattggaagtga 1320  
 Qy 1261 TTATAGGTCAGGTGATTTCAAAATTTCAAAATTTCAAAATTTGATTAATTAATTAATG 1320  
 Db 1321 aaaaatcgcttttaataatgattataatag 1351  
 Qy 1321 AAAAATCGCTTTAATATGATTTATTAATTAATGAG 1351  
 RESULT 2  
 ID V74703 standard; DNA; 2424 BP.  
 AC V74703;  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SEQ ID #392.  
 KW Computer readable medium; vaccine; S. aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.  
 OS Staphylococcus aureus;  
 FH Key Location/Qualifiers  
 FT misc.feature 1141..1200  
 FT /tag= "a  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 PN EP-786519-A2.  
 PD 30-JUL-1997.  
 PR 07-JAN-1997; 100117.  
 PR 05-JAN-1996; US-009861.  
 PA (HMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
 PI Rosen CA;  
 DR WPI; 97-374922/35.  
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 PT stored on computer readable medium and used in the production of  
 PT anti-S. aureus vaccines  
 PS Claim 1; Page 1287-1288; 3271pp; English.  
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S. aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or











PF 04-MAY-1998: U08985.  
PR 14-NOV-1997; US-066009.  
PR 06-MAY-1997; US-044031.  
PR 16-MAY-1997; US-046655.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Dillon PJ, Kunsch CA;  
DR WPI: 99-045171/04.  
PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
PT infection.  
PS Claim 1: Page 1327-1330; 2084pp; English.  
CC A computer readable medium has been developed which has recorded on it  
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC X12938 to X13919 represent these nucleotide sequences which are primary  
CC nucleotide sequences, also known as contigs. The computer-based system  
CC can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring  
CC progression of disease, and for identifying agents which can be used to  
CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
CC another related organism, in vivo or in vitro. In particular the  
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
CC can be used in vaccines to prevent or attenuate an Enterococcal  
CC infection.  
SQ Sequence 4956 BP; 1540 A; 927 C; 842 G; 1635 T;  
Query Match 2.4%; Score 32; DB 60; Length 4956;  
Best Local Similarity 100.0%; Pred. No. 7.78e-10;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 200 catttgcggaattaaagttctgcgatgag 231  
|||  
OY 34 CATTTCGCGAATTAAGGTTCTGCGCATGAG 65  
RESULT 6  
ID X13223 standard; DNA; 2883 BP.  
AC X13223;  
DT 19-MAR-1999 (first entry)  
DE Enterococcus faecalis genome contig SRQ ID NO:286.  
KW Enterococcus faecalis; contig; detection; Enterococcal infection;  
KW vaccine; attenuation; computer readable medium; ds.  
OS Enterococcus faecalis.  
PN WO980555-A2.  
PD 12-NOV-1998.  
PF 04-MAY-1998; U08985.  
PR 14-NOV-1997; US-066009.  
PR 06-MAY-1997; US-044031.  
PR 16-MAY-1997; US-046655.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Dillon PJ, Kunsch CA;  
DR WPI: 99-045171/04.  
PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
PT infection.  
PS Claim 1: Page 1314-1315; 2084pp; English.  
CC A computer readable medium has been developed which has recorded on it  
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC X12938 to X13919 represent these nucleotide sequences which are primary  
CC nucleotide sequences, also known as contigs. The computer-based system  
CC can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring  
CC progression of disease, and for identifying agents which can be used to  
CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
CC another related organism, in vivo or in vitro. In particular the  
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
CC can be used in vaccines to prevent or attenuate an Enterococcal  
CC infection.

SQ Sequence 2883 BP; 990 A; 458 C; 548 G; 881 T;  
Query Match 1.5%; Score 20; DB 60; Length 2883;  
Best Local Similarity 100.0%; Pred. No. 3.60e-01;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 424 attcatttaaatgctgt 443  
|||  
CP 1084 ATTCATTTAATAATGCTGT 1065  
RESULT 7  
ID V99652 standard; DNA; 19 BP.  
AC V99652;  
DT 17-MAR-1999 (first entry)  
DE Murc polynucleotides amplifying primer.  
KW Murc gene; UDP-N-acetylmuramate:L-alanine ligase; Murc polypeptide;  
KW bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
KW immunogen; drug; genetic immunisation; PCR primer; ss.  
OS Synthetic.  
PN Staphylococcus aureus.  
PD EP-889123-A2.  
PD 07-JAN-1999.  
PF 26-JUN-1998; 305064.  
PR 03-JUL-1997; US-052720.  
PA (SMIR ) SMITHKLINE BEECHAM CORP.  
PA (SMIR ) SMITHKLINE BEECHAM PLC.  
PI Burnham MKR, Wallis NG;  
DR WPI: 99-062655/06.  
PT New isolated Murc polypeptide from Staphylococcus aureus and related  
PT nucleic acid - useful in diagnosis, treatment and prevention of  
PT bacterial infections  
PT Disclosure; Page 13; 39pp; English.  
PS The invention relates to a UDP-N-acetylmuramate:L-alanine ligase (Murc  
CC polypeptide) encoded by the S. aureus Murc gene. Host cells containing  
CC an expression system comprising the Murc gene can be used for the  
CC recombinant production of the polypeptide. Agonists or the Murc  
CC polypeptide are used to treat conditions requiring increased activity or  
CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or  
CC competitive polypeptide are useful for inhibiting the polypeptide e.g.  
CC bacterial (especially S. aureus) infections. They are also useful against  
CC Helicobacter pylori infections and related cancers, ulcers and gastritis.  
CC The antibacterial agents are useful to treat in-dwelling devices for  
CC infection prevention or generally as wound treatments to prevent adhesion  
CC of bacteria to matrix proteins. The Murc polypeptide is also useful for  
CC diagnosis or prognosis (a susceptibility to) disease, for raising  
CC antibodies, to identify modulators or specific receptors, in rational  
CC drug design and as an immunogen for vaccines. The Murc gene sequences are  
CC useful in antisense/ribozyme therapeutics; to detect mutant Murc gene;  
CC for chromosomal mapping; to determine bacterial serotype; and for genetic  
CC immunisation. The present sequence represents a primer used for the PCR  
CC amplification of the Murc polynucleotides.  
CC sequence (Murc ORF) of the Murc gene.  
SQ Sequence 19 BP; 10 A; 1 C; 3 G; 5 T;  
Query Match 1.4%; Score 19; DB 60; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.57e+00;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 gttacaataattaagaag 19  
|||  
OY 865 GTTACAATAATTAAGAAG 883  
RESULT 8  
ID V99651 standard; DNA; 19 BP.  
AC V99651;  
DT 17-MAR-1999 (first entry)  
DE Murc polynucleotides amplifying primer.  
KW Murc gene; UDP-N-acetylmuramate:L-alanine ligase; Murc polypeptide;  
KW bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
KW immunogen; drug; genetic immunisation; PCR primer; ss.  
OS Synthetic.



OS Staphylococcus aureus.  
 PN EP-889123-A2.  
 PD 26-JAN-1999.  
 PF 26-JUN-1998: 305064.  
 PR 03-JUL-1997: US-052720.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PI (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Burnham MKR, Wallis NG;  
 PI WPI: 99-062655/06.  
 PT New isolated Murc polypeptide from Staphylococcus aureus and related  
 PT nucleic acid - useful in diagnosis, treatment and prevention of  
 PT bacterial infections  
 PS Disclosure: Page 13: 39pp; English.  
 CC The invention relates to a UDP-N-acetylmuramate:L-alanine ligase (Murc  
 CC polypeptide) encoded by the S. aureus Murc gene. Host cells containing  
 CC an expression system comprising the Murc gene can be used for the  
 CC recombinant production of the polypeptide. Agonists or the Murc  
 CC polypeptide are used to treat conditions requiring increased activity or  
 CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or  
 CC competitive polypeptide are useful for inhibiting the polypeptide e.g.  
 CC bacterial (especially S. aureus) infections. They are also useful against  
 CC Helicobacter pylori infections and related cancers, ulcers and gastritis.  
 CC The antibacterial agents are useful to treat in-dwelling devices for  
 CC infection prevention or generally as wound treatments to prevent adhesion  
 CC of bacteria to matrix proteins. The Murc polypeptide is also useful for  
 CC diagnosing or prognosing a (susceptibility to) disease, for raising  
 CC antibodies, to identify modulators or specific receptors; in rational  
 CC drug design and as an immunogen for vaccines. The Murc gene sequences are  
 CC useful in antisense/ribozyme therapeutics; to detect mutant Murc gene;  
 CC for chromosomal mapping; to determine bacterial serotype; and for genetic  
 CC immunisation. The present sequence represents a primer used for the PCR  
 CC amplification of the Murc ORF of the Murc gene.  
 CC Sequence 19 BP: 6 A; 4 C; 3 G; 6 T;  
 SQ

Query Match 1.4%; Score 19; DB 60; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 1.57e+00;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 cttcataatgaacgatgc 19  
 ||||||||||||||||  
 Cp 1216 CTTCAATATGACGATGC 1198

RESULT 9  
 ID V27381 standard; DNA: 1267 BP.  
 AC V27381:  
 DE 02-OCT-1998 (first entry)  
 DE Streptococcus pneumoniae SP0070 nucleotide.  
 KM Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
 KM detection; pneumonia; otitis media; meningitis; ss.  
 OS Streptococcus pneumoniae.  
 FH Key Location/Qualifiers  
 FT CDS 2..1267  
 FT /\*tag= a  
 FT /\*product= "SP0070"  
 FT /note= "no stop codon given"  
 PN WO9818930-A2.  
 PD 07-MAY-1998.  
 PF 30-OCT-1997: U19422.  
 PR 31-OCT-1996: US-029960.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;  
 PI WPI: 98-272224/24.  
 DR P-PSDB: W55120.  
 PT Nucleic acid encoding antigenic peptide(s) from Streptococcus  
 PT pneumoniae - or their epitope-containing fragments, useful in  
 PT protective or therapeutic vaccines, and for diagnosis  
 PS Claim 1; Page 72; 118pp; English.  
 CC The present sequence encodes a protein from Streptococcus pneumoniae.  
 CC can be useful in vaccines for inducing protective antibodies against  
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.

CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
 CC are used to detect Streptococcus infection (by usual hybridisation or  
 CC amplification methods), also for isolating Streptococcus genes or their  
 CC allelic variants. The protein can be used similarly to detect specific  
 CC antibodies in standard immunoassays, especially for diagnosing or  
 CC monitoring infections. Antibodies which bind the protein are used to  
 CC detect corresponding antigens, to purify the protein and for passive  
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
 CC (especially 10-300) mu g/ml per dose.  
 CC Sequence 1267 BP: 365 A; 269 C; 261 G; 372 T;  
 SQ

Query Match 1.4%; Score 19; DB 46; Length 1267;  
 Best Local Similarity 100.0%; Pred. No. 1.57e+00;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 824 ggtgtaacgctttca 842  
 ||||||||||||||||  
 Qy 901 GGTGTTAAACGCTTTCA 919

RESULT 10  
 ID V43027 standard; DNA: 1825 BP.  
 AC V43027:  
 DE 09-NOV-1998 (first entry)  
 DE Streptococcus pneumoniae polypeptide coding region.  
 DE Streptococcus pneumoniae polypeptide; ORF; open reading frame; bacterial;  
 KM Streptococcus pneumoniae; bacteremia; diagnosis; prophylaxis; ds.  
 OS Streptococcus pneumoniae.  
 FH Key Location/Qualifiers  
 FT CDS complement (256..423)  
 FT /\*tag= a  
 FT /\*note= "polypeptide"  
 FT complement (731..868)  
 FT /\*tag= b  
 FT /note= "polypeptide"  
 PN WO9823631-A1.  
 PD 04-JUN-1998.  
 PF 24-NOV-1997: U21976.  
 PR 27-NOV-1996: US-031879.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO,  
 PI Reid RH, Zarfos PN;  
 PI WPI: 98-322654/28.  
 DR P-PSDB: W62754; W63755.  
 PT Streptococcus pneumoniae polynucleotides - useful for developing  
 PT products for diagnosis, prevention and treatment of infections e.g.  
 PT pneumonia, bacteremia, meningitis or endocarditis  
 PS Claim 1; Page 151-152; 181pp; English.  
 CC The sequence is that of a Streptococcus polypeptide coding region.  
 CC The polypeptide can potentially be used for the diagnosis and  
 CC prevention of bacterial infections, especially SP infection.  
 CC It may be used for the treatment of diseases such as otitis media,  
 CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural  
 CC empyema, endocarditis or infection of the cerebrospinal fluid.  
 CC Sequence 1825 BP: 550 A; 385 C; 366 G; 524 T;  
 SQ

Query Match 1.4%; Score 19; DB 47; Length 1825;  
 Best Local Similarity 100.0%; Pred. No. 1.57e+00;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1160 tgaacgacgtttacacc 1178  
 ||||||||||||||||  
 Cp 919 TGAACGACGTTTAAACCC 901

RESULT 11  
 ID V52194 standard; DNA: 11864 BP.  
 AC V52194:  
 DE 23-OCT-1998 (first entry)  
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:61.  
 DE Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
 KW Streptococcus pneumoniae



KW Computer readable medium; vaccine; pharmaceutical composition; ds.  
 OS Streptococcus pneumoniae.  
 PN WO9818931-A2.  
 PD 07-MAY-1998.  
 PF 30-OCT-1997; U19588.  
 PR 31-OCT-1996; US-009960.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,  
 DR Kunsch CA, Rosen CA;  
 PI WPI: 98-272225/24.  
 PT Computer-readable medium with recorded Streptococcus pneumoniae  
 PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus  
 PT pneumoniae  
 PS Claim 1; Page 524-530; 1409pp; English.  
 CC The present invention describes a computer readable medium which has  
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
 CC on it, or a representative fragment or a sequence at least 95% identical  
 CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
 CC pneumoniae. The present invention also describes an isolated nucleic acid  
 CC molecule encoding a homologue of any of the fragments of the S.pneumoniae  
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 CC by a process comprising: (a) screening a genomic DNA library using as a  
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
 CC to 391, identifying members of the library which contain sequences  
 CC that hybridize to the target sequence and isolating the nucleic acid  
 CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
 CC from an organism, amplifying nucleic acid molecules whose nucleotide  
 CC sequence is homologous to amplification primers derived from the  
 CC fragment of the S. pneumoniae genome to prime the amplification and  
 CC isolating the amplified sequences. The computer readable medium can be  
 CC used in a computer-based system for identifying fragments of the  
 CC S. pneumoniae genome of commercial importance, or expression modulating  
 CC fragments of the S. pneumoniae genome. Products from the present  
 CC invention can be used in diagnosis kits and assays, and pharmaceutical  
 CC compositions and vaccines for S. pneumoniae.  
 CC Sequence 11864 BP; 3440 A; 2668 C; 2183 G; 3573 T;  
 SQ

Query Match 1.48; Score 19; DB 47; Length 11864;  
 Best Local Similarity 100.0%; Pred. No. 1.57e+00;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 855 tgaacgacgtttaacacc 873  
 ||||||||||||||||||  
 Cp 919 TGAACGACGTTTAAACACC 901

RESULT 12  
 ID 036987 standard; DNA; 67 BP.  
 AC 036987;  
 DT 08-JUN-1993 (first entry)  
 DE VLI Ox VL 5'-end PCR primer.  
 KW Spacer peptide; secretase; single chain; fusion protein; antibody;  
 KM scab; recombinant; polymerase chain reaction; ss.  
 OS Synthetic.  
 PN FI9103434-A.  
 PD WO9302198-A.  
 PD 17-JAN-1992.  
 PF 16-JUL-1991; 913434.  
 PR 16-JUL-1990; US-552751.  
 PA (TERE-) TECH RES CENT FINLAND.  
 PI Alfthan K, Knowles JKC, Laukkanen ML, Sizmann D, Takkinen K, Teeri TT;  
 DR WPI: 92-134225/17.  
 PT Prod. of single chain fusion protein, pref. antibody - comprises  
 PT transforming host cells, e.g. E. coli with expression constructs  
 PT composed of proteins or domains, linked by spacer peptide(s)  
 PS Example; Page 28; 56pp; English.  
 CC The sequence is that of the PCR primer VLI, which is used in the  
 CC modification, for in-frame fusions with the hinge coding region,  
 CC of the 5'-end of Ox VL cDNA.  
 CC Sequence 67 BP; 16 A; 22 C; 11 G; 18 T;

Query Match 1.38; Score 17; DB 6; Length 67;  
 Best Local Similarity 100.0%; Pred. No. 2.60e+01;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 37 cattatcaattgtttc 53  
 ||||||||||||||||||  
 Cp 1249 CATTATCAATGTGTTCT 1233

RESULT 13  
 ID V75600 standard; DNA; 154 BP.  
 AC V75600;  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SEQ ID #1289.  
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 OS Staphylococcus aureus.  
 PN EP-786519-A2.  
 PD 30-JUL-1997.  
 PF 07-JAN-1997; 100117.  
 PR 05-JAN-1996; US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
 PI Rosen CA;  
 PI WPI: 97-374922/35.  
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 PT stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines  
 PS Claim 1; Page 1907; 3277pp; English.  
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.  
 CC Sequence 154 BP; 42 A; 32 C; 25 G; 54 T;

Query Match 1.38; Score 18; DB 56; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 6.54e+00;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 23 ggcattatcatgttat 40  
 ||||||||||||||||||  
 QY 1327 GCGTTTAAATGTTTAT 1344

RESULT 14  
 ID V89109 standard; cDNA; 458 BP.  
 AC V89109;  
 DT 15-FEB-1999 (first entry)  
 DE EST clone BZ187.  
 KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
 KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;  
 KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
 OS Homo sapiens.  
 PN WO9845436-A2.  
 PD 15-OCT-1998.  
 PF 10-APR-1998; U06955.



PR 10-APR-1997: US-838821.  
 PA (GENEX) GENETICS INST INC.  
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
 PI Racine LA, Spaulding V, Treacy M;  
 DR WPI: 99-070077/06.  
 PT New polynucleotides encoding human secreted proteins - derived from  
 e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PI ovary, pituitary, retina and colon cDNA libraries.  
 PS Claim 1: Page 113: 618pp; English.  
 CC The present sequence represents a human expressed sequence tag (EST).  
 CC The polynucleotide, which is a secreted EST, and the encoded protein  
 CC are predicted to have useful biological activities which would make  
 CC them suitable for treating, preventing or ameliorating medical  
 CC conditions in humans and animals, although no supporting data is  
 CC given. Suggested activities include nutritional activity, immune  
 CC stimulating or suppressing activity, haematopoiesis regulating  
 CC activity, tissue growth activity, activin/inhibin activity,  
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
 CC activity, receptor/ligand activity, anti-inflammatory activity,  
 CC cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The polynucleotide may also be useful for gene therapy.  
 SQ Sequence 458 BP; 152 A; 70 C; 74 G; 162 T;

Query Match 1.3%; Score 17; DB 53; Length 458;  
 Best Local Similarity 100.0%; Pred. No. 2.60e+01;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 334 atctataatgtatla 350  
 ||||||||||||||||  
 QY 1217 ATCTATTAATGATTA 1233

RESULT 15  
 ID 059532 standard; cDNA; 474 BP.  
 AC 059532;  
 DT 16-MAR-1994 (first entry)  
 DE Human brain Expressed Sequence Tag EST00527.  
 KW Gene transcription product; genetic markers; tagging; in vivo;  
 KW transcription; mapping; locations; chromosomes; chromosomal; ss.  
 OS Homo sapiens.  
 PN W09316178-A.  
 PD 19-AUG-1993.  
 PF 12-FEB-1993; U01294.  
 PF 12-FEB-1992; US-837195.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 PI Adams MD, Moreno RF, Venter CJ;  
 DR WPI: 93-272882/34.  
 PT Enriched oligonucleotides and corresp. sequences - used as  
 PT markers for human genes transcribed in-vivo, facilitate tagging  
 PT of most human genes  
 PS Example 4: Page 190: 500pp; English.  
 CC The Expressed Sequence Tag was isolated from a human brain cDNA  
 CC library as part of a large set of ESTs which can be used as markers  
 CC for human genes transcribed in vivo. They can be used to facilitate  
 CC tagging of most human genes, for mapping locations of expressed genes  
 CC on chromosomes, for individual or forensic identification, for mapping  
 CC locations of disease-associated genes, for identification of tissue  
 CC type, and for prepn. of antisense sequences, probes and constructs.  
 CC EST00527 has a "poor" coding probability as evaluated using the  
 CC coding-region prediction program CRM. See also Q59041-Q061440.  
 SQ Sequence 474 BP; 155 A; 88 C; 75 G; 153 T;

Query Match 1.3%; Score 17; DB 8; Length 474;  
 Best Local Similarity 100.0%; Pred. No. 2.60e+01;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 372 cattttaatgaattt 388  
 ||||||||||||||||  
 QY 1070 CATTTTAATGATTT 1086







CC LIBRARY: BRSTNOT03  
CC CLONE: 640699  
SQ SEQUENCE 853 BP; 233 A; 187 C; 162 G; 271 T; 0 OTHER.

Query Match 1.3%; Score 17; DB 3; Length 853;  
Best Local Similarity 100.0%; Pred. No. 2.19e+00;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 685 ATCTATTATGATTA 701  
1217 ATCTATTATGATTA 1233

RESULT 2  
ID US-08-474-499-7 STANDARD; DNA; UNC; 1791 BP.  
AC xxxxxx

DE Sequence 7, Application US/08474499  
CC Sequence 7, Application US/08474499  
CC Patent No. 5693776  
CC GENERAL INFORMATION:  
CC APPLICANT: Anderson, Burt E.  
CC TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
CC TITLE OF INVENTION: DIAGNOSING ROCHALIMAEA HENSELAE  
CC TITLE OF INVENTION: AND ROCHALIMAEA QUINTANA INFECTION  
CC NUMBER OF SEQUENCES: 10  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
CC STREET: 127 Peachtree Street, Suite 1200  
CC City: Atlanta  
CC STATE: Georgia  
CC COUNTRY: USA  
CC ZIP: 30303  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/474,499  
CC FILING DATE: 07-JUN-1995  
CC CLASSIFICATION: 536  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/245,294  
CC FILING DATE: 18-MAY-1994  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Spratt, Gwendolyn D.  
CC REGISTRATION NUMBER: 36,016  
CC REFERENCE/DOCKET NUMBER: 1414.612  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 404/688-0770  
CC TELEFAX: 404/688-9880  
CC INFORMATION FOR SEQ ID NO: 7:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1791 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 141..1649  
SQ SEQUENCE 1791 BP; 555 A; 263 C; 408 G; 565 T; 0 OTHER.

Query Match 1.3%; Score 17; DB 2; Length 1791;  
Best Local Similarity 100.0%; Pred. No. 2.19e+00;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1345 GCTCAAAATATTCAAAT 1361  
712 GCTCAAAATATTCAAAT 728

RESULT 3  
ID US-08-245-294-7 STANDARD; DNA; UNC; 1791 BP.  
AC xxxxxx

DE Sequence 7, Application US/08245294  
CC Sequence 7, Application US/08245294  
CC Patent No. 5644047  
CC GENERAL INFORMATION:  
CC APPLICANT: Anderson, Burt E.  
CC TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
CC TITLE OF INVENTION: DIAGNOSING  
CC TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA  
CC NUMBER OF SEQUENCES: 10  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
CC STREET: 127 Peachtree Street, Suite 1200  
CC City: Atlanta  
CC STATE: Georgia  
CC COUNTRY: USA  
CC ZIP: 30303  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: IBM PC compatible  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/245,294  
CC FILING DATE:  
CC CLASSIFICATION: 424  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Spratt, Gwendolyn D.  
CC REGISTRATION NUMBER: 36,016  
CC REFERENCE/DOCKET NUMBER: 1414.612  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 404/688-0770  
CC TELEFAX: 404/688-9880  
CC INFORMATION FOR SEQ ID NO: 7:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1791 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 141..1649  
SQ SEQUENCE 1791 BP; 555 A; 263 C; 408 G; 565 T; 0 OTHER.

Query Match 1.3%; Score 17; DB 1; Length 1791;  
Best Local Similarity 100.0%; Pred. No. 2.19e+00;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1345 GCTCAAAATATTCAAAT 1361  
712 GCTCAAAATATTCAAAT 728

RESULT 4  
ID PCT-US95-06211-7 STANDARD; DNA; UNC; 1791 BP.  
AC xxxxxx

DE Sequence 7, Application PC/TUS9506211  
CC Sequence 7, Application PC/TUS9506211  
CC GENERAL INFORMATION:  
CC APPLICANT:  
CC TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
CC TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA INFECTION  
CC NUMBER OF SEQUENCES: 10  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: NEEDLE & ROSENBERG, P.C.



CC STREET: 127 Peachtree Street, Suite 1200  
CC City: Atlanta  
CC State: Georgia  
CC Country: USA  
CC ZIP: 30303  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC PCT/US95/06211  
CC APPLICATION NUMBER: US 08/245,294  
CC FILING DATE: 18 MAY 1994  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Spratt, Gwendolyn D.  
CC REGISTRATION NUMBER: 36,016  
CC REFERENCE/DOCKET NUMBER: 1414.6121  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 404/688-0770  
CC TELEFAX: 404/688-9880  
CC INFORMATION FOR SEQ ID NO: 7:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1791 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 141..1649  
CC SEQUENCE 1791 BP; 555 A; 263 C; 408 G; 565 T; 0 OTHER.

Query Match 1.3%; Score 17; DB 4; Length 1791;  
Best Local Similarity 100.0%; Pred. No. 2.19e+00;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1345 GCTCAAAATATTCAAAT 1361  
|||||  
QY 712 GCTCAAAATATTCAAAT 728

RESULT 5  
ID US-08-307-279A-7 STANDARD; DNA; UNC; 1791 BP.  
AC xxxxxx  
DT  
DE Sequence 7, Application US/08307279A  
CC Sequence 7, Application US/08307279A  
CC Patent No. 5736347  
CC GENERAL INFORMATION:  
CC APPLICANT: Anderson, Burt E.  
CC APPLICANT: Regnery, Russell L  
CC TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae  
CC TITLE OF INVENTION: and Methods and Compositions for Diagnosing Rochalimaea  
CC NUMBER OF SEQUENCES: 14  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
CC STREET: 127 Peachtree Street, N.E., Suite 1200  
CC City: Atlanta  
CC State: Georgia  
CC Country: USA  
CC ZIP: 30303  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/307,279A  
CC FILING DATE:

CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Spratt, Gwendolyn D.  
CC REGISTRATION NUMBER: 36,016  
CC REFERENCE/DOCKET NUMBER: 1414.624  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (404) 688-0770  
CC TELEFAX: (404) 688-9880  
CC INFORMATION FOR SEQ ID NO: 7:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1791 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 141..1652  
CC SEQUENCE 1791 BP; 555 A; 263 C; 408 G; 565 T; 0 OTHER.

Query Match 1.3%; Score 17; DB 2; Length 1791;  
Best Local Similarity 100.0%; Pred. No. 2.19e+00;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1345 GCTCAAAATATTCAAAT 1361  
|||||  
QY 712 GCTCAAAATATTCAAAT 728

RESULT 6  
ID US-08-421-661-5 STANDARD; DNA; UNC; 2465 BP.  
AC xxxxxx  
DT  
DE Sequence 5, Application US/08421661  
CC Sequence 5, Application US/08421661  
CC Patent No. 5807993  
CC GENERAL INFORMATION:  
CC APPLICANT: French, Cynthia K  
CC APPLICANT: Yamamoto, Karen K  
CC APPLICANT: Chow, Phoebe M  
CC APPLICANT: Alido, Nemesias T  
CC TITLE OF INVENTION: Antigen Related to Inflammatory Diseases  
CC NUMBER OF SEQUENCES: 6  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: David J. Oldenkamp  
CC STREET: 2029 Century Park East, Suite 3800  
CC City: Los Angeles  
CC State: California  
CC Country: U.S.  
CC ZIP: 90067  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/421,661  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Oldenkamp, David J  
CC REGISTRATION NUMBER: 29,421  
CC REFERENCE/DOCKET NUMBER: 109-067  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 310-788-5000  
CC TELEFAX: 310-277-1297  
CC INFORMATION FOR SEQ ID NO: 5:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 2465 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: CDNA



CC HYPOTHETICAL: NO  
CC ANTI-SENSE: NO  
CC FRAGMENT TYPE: N-terminal  
CC ORIGINAL SOURCE:  
CC ORGANISM: Homo sapiens  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 61..2054  
SQ SEQUENCE 2465 BP; 836 A; 488 C; 541 G; 580 T; 0 OTHER.

Query Match 1.3%; Score 17; DB 3; Length 2465;  
Best Local Similarity 100.0%; Pred. No. 2.19e+00;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 422 ACATAAAGAGATATG 438  
|||||  
Qy 191 ACATAAAGAGATATG 207

RESULT 7  
ID US-08-967-101-17 STANDARD; DNA; UNC; 373 BP.  
AC xxxxxx  
DT  
DE Sequence 17, Application US/08967101  
CC Sequence 17, Application US/08967101  
CC Patent No. 5840540  
CC GENERAL INFORMATION:  
CC APPLICANT: ST. GEORGE-HYSLOP, PETER H  
CC APPLICANT: ROMMENS, JOHANNA M  
CC APPLICANT: FRASER, PAUL E  
CC TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
CC TITLE OF INVENTION: TO ALZHEIMER'S DISEASE  
CC NUMBER OF SEQUENCES: 183  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
CC STREET: High Street Tower - 125 High Street  
CC CITY: Boston  
CC STATE: Massachusetts  
CC COUNTRY: U.S.A.  
CC ZIP: 02110  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/967,101  
CC FILING DATE: 10-NOV-1997  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/592,541  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Pletcher, Edmund R.  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 248-7100  
CC TELEFAX: (617) 248-7100  
CC INFORMATION FOR SEQ ID NO: 17:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 373 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
SQ SEQUENCE 373 BP; 117 A; 64 C; 56 G; 115 T; 21 OTHER.

Query Match 1.2%; Score 16; DB 3; Length 373;  
Best Local Similarity 100.0%; Pred. No. 1.05e+01;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 325 ATGTAATATTTTAT 340  
|||||  
Cp 178 ATGTAATATTTTAT 163

RESULT 8  
ID US-08-474-542A-20 STANDARD; DNA; UNC; 415 BP.  
AC xxxxxx  
DT  
DE Sequence 20, Application US/08474542A  
CC Sequence 20, Application US/08474542A  
CC Patent No. 5527898  
CC GENERAL INFORMATION:  
CC APPLICANT: Bauer, Heidi M.  
CC APPLICANT: Gravitt, Patti E.  
CC APPLICANT: Greer, Catherine E.  
CC APPLICANT: Imptaim, Chaka C.  
CC APPLICANT: Manos, M. Michele  
CC APPLICANT: Resnick, Robert M.  
CC TITLE OF INVENTION: Detection of Human Papillomavirus by the  
CC TITLE OF INVENTION: Polymerase Chain Reaction  
CC NUMBER OF SEQUENCES: 298  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Hoffmann-La Roche Inc.  
CC STREET: 340 Kingsland Street  
CC CITY: Nutley  
CC STATE: New Jersey  
CC COUNTRY: U.S.A.  
CC ZIP: 07110  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/474,542A  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Petry, Douglas A.  
CC REGISTRATION NUMBER: 35,321  
CC REFERENCE/DOCKET NUMBER: 9234  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (510) 814-2974  
CC TELEFAX: (510) 814-2977  
CC INFORMATION FOR SEQ ID NO: 20:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 415 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
SQ SEQUENCE 415 BP; 127 A; 72 C; 74 G; 142 T; 0 OTHER.

Query Match 1.2%; Score 16; DB 1; Length 415;  
Best Local Similarity 100.0%; Pred. No. 1.05e+01;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 58 TTAAACATTATGTGCC 73  
|||||  
Cp 612 TTAAACATTATGTGCC 597

RESULT 9  
ID US-08-457-648-20 STANDARD; DNA; UNC; 415 BP.  
AC xxxxxx  
DT  
DE Sequence 20, Application US/08457648  
CC Sequence 20, Application US/08457648  
CC Patent No. 5639871  
CC GENERAL INFORMATION:  
CC APPLICANT: Bauer, Heidi M.  
CC APPLICANT: Gravitt, Patti E.  
CC APPLICANT: Greer, Catherine E.  
CC APPLICANT: Imptaim, Chaka C.  
CC APPLICANT: Manos, M. Michele



CC APPLICANT: Resnick, Robert M.  
CC TITLE OF INVENTION: Detection of Human Papillomavirus by the  
CC TITLE OF INVENTION: Polymerase Chain Reaction  
CC NUMBER OF SEQUENCES: 298  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Hoffmann-La Roche Inc.  
CC STREET: 340 Kingsland Street  
CC CITY: Nutley  
CC STATE: New Jersey  
CC COUNTRY: U.S.A.  
CC ZIP: 07110  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/457,648  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Petry, Douglas A.  
CC REGISTRATION NUMBER: 35,321  
CC REFERENCE/DOCKET NUMBER: 9205  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (510) 814-2974  
CC TELEFAX: (510) 814-2974  
CC INFORMATION FOR SEQ ID NO: 20:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 415 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
SQ SEQUENCE 415 BP; 127 A; 72 C; 74 G; 142 T; 0 OTHER.

Query Match 1.28; Score 16; DB 1; Length 415;  
Best Local Similarity 100.08; Pred. No. 1.05e+01;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 58 TTAACTATTGTGCC 73  
CP 612 TTAACTATTGTGCC 597

RESULT 10  
ID US-08-474-542A-288 STANDARD; DNA; UNC; 467 BP.  
AC xxxxxx  
DE Sequence 288, Application US/08474542A  
DE Sequence 288, Application US/08474542A  
CC Patent No. 5527898  
CC GENERAL INFORMATION:  
CC APPLICANT: Bauer, Heidi M.  
CC APPLICANT: Gravitt, Patti E.  
CC APPLICANT: Greer, Catherine E.  
CC APPLICANT: Impraal, Chaka C.  
CC APPLICANT: Manos, M. Michele  
CC APPLICANT: Resnick, Robert M.  
CC TITLE OF INVENTION: Detection of Human Papillomavirus by the  
CC TITLE OF INVENTION: Polymerase Chain Reaction  
CC NUMBER OF SEQUENCES: 298  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Hoffmann-La Roche Inc.  
CC STREET: 340 Kingsland Street  
CC CITY: Nutley  
CC STATE: New Jersey  
CC COUNTRY: U.S.A.  
CC ZIP: 07110  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/474,542A  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Petry, Douglas A.  
CC REGISTRATION NUMBER: 35,321  
CC REFERENCE/DOCKET NUMBER: 9234  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (510) 814-2974  
CC TELEFAX: (510) 814-2977  
CC INFORMATION FOR SEQ ID NO: 288:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 467 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
SQ SEQUENCE 467 BP; 140 A; 84 C; 91 G; 152 T; 0 OTHER.

Query Match 1.28; Score 16; DB 1; Length 467;  
Best Local Similarity 100.08; Pred. No. 1.05e+01;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 TTAACTATTGTGCC 99  
CP 612 TTAACTATTGTGCC 597

RESULT 11  
ID US-08-457-648-288 STANDARD; DNA; UNC; 467 BP.  
AC xxxxxx  
DE Sequence 288, Application US/08457648  
DE Sequence 288, Application US/08457648  
CC Patent No. 5639871  
CC GENERAL INFORMATION:  
CC APPLICANT: Bauer, Heidi M.  
CC APPLICANT: Gravitt, Patti E.  
CC APPLICANT: Greer, Catherine E.  
CC APPLICANT: Impraal, Chaka C.  
CC APPLICANT: Manos, M. Michele  
CC APPLICANT: Resnick, Robert M.  
CC TITLE OF INVENTION: Detection of Human Papillomavirus by the  
CC TITLE OF INVENTION: Polymerase Chain Reaction  
CC NUMBER OF SEQUENCES: 298  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Hoffmann-La Roche Inc.  
CC STREET: 340 Kingsland Street  
CC CITY: Nutley  
CC STATE: New Jersey  
CC COUNTRY: U.S.A.  
CC ZIP: 07110  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/457,648  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Petry, Douglas A.  
CC REGISTRATION NUMBER: 35,321  
CC REFERENCE/DOCKET NUMBER: 9205  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (510) 814-2974  
CC TELEFAX: (510) 814-2977  
CC INFORMATION FOR SEQ ID NO: 288:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 467 base pairs



CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
SQ SEQUENCE 467 BP; 140 A; 84 C; 91 G; 152 T; 0 OTHER.

Query Match 1.2%; Score 16; DB 1; Length 467;  
Best Local Similarity 100.0%; Pred. No. 1.05e+01;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 84 TTATACATTATGTGCC 99  
CP 612 TTATACATTATGTGCC 597

RESULT 12  
ID US-08-307-499-51 STANDARD; DNA; UNC; 810 BP.  
AC xxxxxx  
DE Sequence 51, Application US/08307499  
CC Sequence 51, Application US/08307499  
CC Patent No. 5651972  
CC GENERAL INFORMATION:  
CC APPLICANT: Moyer, Richard W.  
CC APPLICANT: Vi uela, Eladio  
CC APPLICANT: Gibbs, E.P.J.  
CC TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a  
CC TITLE OF INVENTION: Live Vaccine Vector  
CC NUMBER OF SEQUENCES: 60  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: David R. Saliwanchik  
CC STREET: 2421 N.W. 41st Street, Suite A-1  
CC CITY: Gainesville  
CC STATE: Florida  
CC COUNTRY: U.S.A.  
CC ZIP: 32606  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/307,499  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/908,241  
CC FILING DATE: 1-JUL-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/908,630  
CC FILING DATE: 29-JUN-1992  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/342,212  
CC FILING DATE: 21-APR-1992  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Saliwanchik, David R.  
CC REGISTRATION NUMBER: 31,794  
CC REFERENCE/DOCKET NUMBER: UF35.1.FWCC1  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 904-375-8100  
CC TELEFAX: 904-372-5800  
CC INFORMATION FOR SEQ ID NO: 51:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 810 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: double  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: DNA (genomic)  
SQ SEQUENCE 810 BP; 260 A; 127 C; 133 G; 290 T; 0 OTHER.  
Query Match 1.2%; Score 16; DB 1; Length 810;

Best Local Similarity 100.0%; Pred. No. 1.05e+01;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 123 TATATAGTTTATATA 138  
QY 1333 TATATAGTTTATATA 1348

RESULT 13  
ID US-08-413-118-117 STANDARD; DNA; UNC; 1760 BP.  
AC xxxxxx  
DE Sequence 117, Application US/08413118  
CC Sequence 117, Application US/08413118  
CC Patent No. 5688920  
CC GENERAL INFORMATION:  
CC APPLICANT: PAOLETTI, ENZO  
CC APPLICANT: LIMBACH, KEITH J.  
CC TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
CC TITLE OF INVENTION: CANINE HERPESVIRUS gB, gC, AND gD AND USES THEREFOR  
CC NUMBER OF SEQUENCES: 128  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.  
CC STREET: 530 FIFTH AVENUE, 25TH FLOOR  
CC CITY: NEW YORK  
CC STATE: NEW YORK  
CC COUNTRY: UNITED STATES OF AMERICA  
CC ZIP: 10036  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/413,118  
CC FILING DATE: 29-MAR-1995  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/220,151  
CC FILING DATE: 30-MAR-1994  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: PROWSE, WILLIAM S.  
CC REGISTRATION NUMBER: 25,506  
CC REFERENCE/DOCKET NUMBER: 454310-2670  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 840-3333  
CC TELEFAX: (212) 840-0712  
CC INFORMATION FOR SEQ ID NO: 117:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1760 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: CDNA  
SQ SEQUENCE 1760 BP; 695 A; 243 C; 230 G; 592 T; 0 OTHER.  
Query Match 1.2%; Score 16; DB 2; Length 1760;  
Best Local Similarity 100.0%; Pred. No. 1.05e+01;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 755 TTATATAGTTTATTA 770  
QY 288 TTATATAGTTTATTA 303  
RESULT 14  
ID US-07-803-622E-3 STANDARD; DNA; UNC; 1821 BP.  
AC xxxxxx  
DE Sequence 3, Application US/07803622E  
CC Sequence 3, Application US/07803622E  
CC Patent No. 5525497  
CC GENERAL INFORMATION:



```

CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/070,165F
CC FILING DATE:
CC CLASSIFICATION: 435
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (614)-293-8093
CC TELEFAX: (614)-293-5631
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1839 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC ORIGINAL SOURCE:
CC ORGANISM: NO. 5750365ophthalmus viridescens
CC DEVELOPMENTAL STAGE: Adult
CC TISSUE TYPE: Regenerating forelimb blastema
CC CELL TYPE: Mesenchyme and Epithelium
CC IMMEDIATE SOURCE:
CC LIBRARY: lambda gt11
CC CLONE: Mj3-1
CC POSITION IN GENOME:
CC UNITS: bp
CC FEATURE:
CC NAME/KEY: CDS
CC LOCATION: 1..1134
CC FEATURE:
CC NAME/KEY: 3'UTR
CC LOCATION: 1135..1839
CC PUBLICATION INFORMATION:
CC AUTHORS: Poulin, Matthew L
CC RELEVANT RESIDUES IN SEQ ID NO: 7: FROM 1 TO 1839
SQ SEQUENCE 1839 BP; 502 A; 400 C; 469 G; 468 T; 0 OTHER.

Query Match 1.2%; Score 16; DB 2; Length 1839;
Best Local Similarity 100.0%; Pred. No. 1.05e+01;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Db 306 CAAAAACATCATTAAT 321
|||||
Cp 580 CAAAAACATCATTAAT 565

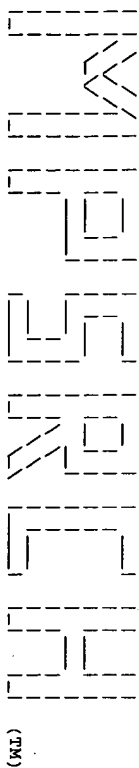
Search completed: Sat Nov 27 14:17:01 1999
Job time : 104 secs.

```



**THIS PAGE BLANK (USPTO)**





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_n n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Sat Nov 27 13:30:35 1999; Maspar time 2278.20 Seconds  
1389.523 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-103-287-1  
Description: (1-1351) from US09103287.seq  
Perfect Score: 1351  
N.A. Sequence: 1 ATGAGTAAAGGAGTTTATAT.....TTAATATGTTTATATATAGAG 1351  
Comp: TACTCATCTCTCAAAATATATA.....AATTATCAATATATATCTC

Scoring table: TABLE Jmetric  
Gap 60

Mismatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb1-est58  
1:em\_est10.2:em\_est11.3:em\_est17.4:em\_est18.5:em\_est12  
6:em\_est19.7:em\_gss1

Database: genbank-est111  
8:gb\_est1.9:gb\_est10.10:gb\_est11.11:gb\_est12.12:gb\_est13  
13:gb\_est14.14:gb\_est15.15:gb\_est16.16:gb\_est17  
17:gb\_est18.18:gb\_est19.19:gb\_est20.20:gb\_est21  
21:gb\_est22.22:gb\_est23.23:gb\_est24.24:gb\_est25  
25:gb\_est26.26:gb\_est27.27:gb\_est28.28:gb\_est29  
29:gb\_est30.30:gb\_est31.31:gb\_est32.32:gb\_est33.33:gb\_est34  
34:gb\_est35.35:gb\_est36.36:gb\_est37.37:gb\_gss1.38:gb\_gss2  
39:gb\_gss3.40:gb\_gss4.41:gb\_gss5.42:gb\_gss6

Statistics: Mean 9.442; Variance 1.331; scale 7.096

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
c 1	20	1.5	244 31	D62295	HUM253D09B Clontech hu	4.91e-06
c 2	20	1.5	351 31	H35671	EST105257 Rat PC-12 ce	4.91e-06
c 3	20	1.5	429 26	AA956076	UI-R-EI-In-b-08-0-UI's	4.91e-06
c 4	20	1.5	431 26	AA956542	UI-R-EI-In-e-12-0-UI's	4.91e-06
c 5	20	1.5	435 21	AA943324	EST198823 Normalized r	4.91e-06
c 6	20	1.5	460 21	AI013962	EST207517 Normalized r	4.91e-06
c 7	20	1.5	571 23	AI176934	EST20540 Normalized r	4.91e-06
c 8	20	1.5	633 26	AI008955	EST203406 Normalized r	4.91e-06
c 9	20	1.5	667 38	B68124	CT1978SK-R-802H9.TP CI	4.91e-06
c 10	19	1.4	196 13	AA480754	SMFCA2133SK Bruglia ma	2.10e-04

c 11	19	1.4	232 39	AO195928	RPc111-66E20.TK RPc111	2.10e-04
c 12	19	1.4	332 40	AO233339	HS-2014.A2.G02.T7 CIT	2.10e-04
c 13	19	1.4	376 36	AA080511	m888e11.r1 Stratagene	2.10e-04
c 14	19	1.4	383 39	AO115573	RPc111-57N20.TK RPc111	2.10e-04
c 15	19	1.4	399 39	AO093108	HS-3020.B1.A12.T7 CIT	2.10e-04
c 16	19	1.4	412 41	AO303001	HS-3113.B2.G07.MR CIT	2.10e-04
c 17	19	1.4	423 39	AO108295	CIT-HSP-2380119.TF CIT	2.10e-04
c 18	19	1.4	477 38	B85058	RPc111-29M5.TV RPc111	2.10e-04
c 19	19	1.4	497 24	AI256065	u194f07.r1 Sugano mous	2.10e-04
c 20	19	1.4	519 36	AA144425	mr16h12.r1 Soares mous	2.10e-04
c 21	19	1.4	545 39	AO139297	HS-3079.A2.D12.MF CIT	2.10e-04
c 22	19	1.4	552 41	AO322849	RPc111-11102.TJ RPc111	2.10e-04
c 23	19	1.4	614 14	C25739	C25739 Dictyostelium d	2.10e-04
c 24	19	1.4	618 42	AO415275	RPc1-11-180011.TV RPc1	2.10e-04
c 25	19	1.4	628 22	AI055437	coeu000401 Cotton Bol	2.10e-04
c 26	19	1.4	631 37	B67375	T24H9TF TAMU Arabidops	2.10e-04
c 27	19	1.4	674 20	C98676	C89676 Dictyostelium d	2.10e-04
c 28	19	1.4	684 38	B54100	CIT-HSP-2014K4.TF CIT	2.10e-04
c 29	18	1.3	154 10	AA262130	zs23b08.r1 NC1.CGAP.GC	7.52e-03
c 30	18	1.3	219 20	AA886233	o121h02.s1 NC1.CGAP.K1	7.52e-03
c 31	18	1.3	273 38	AO067661	HS-2239.B2.A10.MF CIT	7.52e-03
c 32	18	1.3	300 15	C57017	C57017 Yui Kohara unp	7.52e-03
c 33	18	1.3	300 15	C52223	C52223 Yui Kohara unp	7.52e-03
c 34	18	1.3	328 38	AO057313	CIT-HSP-2337K23.TF CIT	7.52e-03
c 35	18	1.3	357 15	AA569728	nt25b09.s1 NC1.CGAP.Pr	7.52e-03
c 36	18	1.3	387 28	AI535161	UI-R-C3-89-C-03-0-UI's	7.52e-03
c 37	18	1.3	388 18	AA825640	o67d04.s1 NC1.CGAP.GC	7.52e-03
c 38	18	1.3	398 32	H83996	y93g09.s1 Soares reti	7.52e-03
c 39	18	1.3	405 21	AA937979	OC08f08.s1 NC1.CGAP.GC	7.52e-03
c 40	18	1.3	436 23	AI145238	UI-R-BT0-qf-C-02-0-UI	7.52e-03
c 41	18	1.3	548 28	AI509035	va78f09.y1 Soares mous	7.52e-03
c 42	18	1.3	568 17	AA540021	LD19344.Sprtime ID Dros	7.52e-03
c 43	18	1.3	582 20	AA891693	EST195496 Normalized r	7.52e-03
c 44	18	1.3	586 28	AI546300	LD47843.Sprtime ID Dros	7.52e-03
c 45	18	1.3	668 41	AO282741	RPc111-89E10.TV RPc111	7.52e-03

## ALIGNMENTS

RESULT	1												
LOCUS	D62295	244 bp	mRNA	EST	29-AUG-1995								
DEFINITION	HUM253D09B Clontech human aorta polyA+ mRNA (#6572)			Homo sapiens									
ACCESSION	CDNA clone GGN-253D09 5', mRNA sequence.												
NID	D62295												
VERSION	9966071												
KEYWORDS	D62295.1	GI:966071											
SOURCE	EST.												
ORGANISM	human.												
	Homo sapiens												
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;												
	Eutheria; Primates; Catarrhini; Homnidae; Homo.												
REFERENCE	1 (bases 1 to 244)												
AUTHORS	Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M., Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shimomura,H., Takachi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y., Maekawa,H., Shin,S. and Nakamura,Y.												
TITLE	Unpublished (1995)												
JOURNAL													
COMMENT	On Apr 14, 1993 this sequence version replaced gi:693074.												

## FEATURES

source  
1. 244  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="13"  
/clone="GEN-253D09"  
/clone\_lib="Clontech human aorta polyA+ mRNA (#6572)"

Contact: Tsutomu Fujiwara  
Otsuka GEN Research Institute  
Otsuka Pharmaceutical Co., Ltd  
463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan  
Tel: 0886-65-2888  
Fax: 0886-37-1035.  
Location/Qualifiers



```

BASE COUNT      81 a      31 c      45 g      85 t      2 others
ORIGIN

Query Match      1.58; Score 20; DB 31; Length 244;
Best Local Similarity 100.0%; Pred. No. 4,91e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      224 AAACATCATTAATATCTTT 243
CP      578 AAACATCATTAATATCTTT 559

RESULT      2
LOCUS      H35671      351 bp      mRNA      EST      02-APR-1998
DEFINITION EST0105257 Rat PC-12 cells, untreated Rattus sp. cDNA clone RPCA40
mRNA sequence.
ACCESSION   H35671
VERSION     9981088
KEYWORDS    H35671.1 GI:981088
SOURCE      EST
ORGANISM    Rattus sp.
            Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 351)
            Lee,N.H., Weinstein,K.G., Kirkness,E.F., Earle-Hughes,J.A.,
            Fuldner,R.A., Marmaras,S., Glodex,A., Gocayne,J.D., Adams,M.D.,
            Kerlavage,A.R., Fraser,C.M. and Venter,J.C.
            Comparative expressed-sequence-tag analysis of differential gene
            expression profiles in PC-12 cells before and after nerve growth
            factor treatment
            Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)
JOURNAL     95396786
MEDLINE     On May 8, 1995 this sequence version replaced gi:800622.
COMMENT     Other ESTs: TC797
            Contact: Lee, NH
            ATCC

The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
For clone availability please contact the TIGR Database
(tdbinfo@cdb.tigr.org) TC (Tentative Consensus) numbers represent
assemblies of ESTs.

FEATURES
source      1..351
            Location/Qualifiers
            1..351
            /organism="Rattus sp."
            /note="vector: phuscript SK-; Site_1: EcoRI; Site_2:
            XhoI; poly(A)+ RNA was purified from untreated PC12 cells
            cultured for 9 days. cDNA was constructed using an
            oligo-dT primer and directionally cloned using the lambda
            Zap II Vector Kit by Stratagene"
            /db_xref="ATCC (Inhost):2000417"
            /db_xref="taxon:10118"
            /clone="RPCA40"
            /clone_1lb="Rat PC-12 cells, untreated"

BASE COUNT      116 a      68 c      80 g      87 t

Query Match      1.58; Score 20; DB 31; Length 351;
Best Local Similarity 100.0%; Pred. No. 4,91e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      22 TATTATAACATATTAAAC 41
CP      1348 TATTATAACATATTAAAC 1329

RESULT      3
LOCUS      AA956076      429 bp      tRNA      EST      07-FEB-1999
DEFINITION UT-R-El-fn-b-08-0-UI.s1 UT-R-El Rattus norvegicus cDNA clone
UT-R-El-fn-b-08-0-UI 3' similar to g112957159|gb|AC004404|AC004404

```

```

sequence. Mus musculus ma08m078, complete sequence [Mus musculus], mRNA
ACCESSION      AA956076
NID            94238470
VERSION        AA956076.1  GI:4238470
KEYWORDS
SOURCE
ORGANISM       Rattus norvegicus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Rattus.
REFERENCE      1 (bases 1 to 429)
               Bonaldo,M.F., Lennon,G. and Soares,M.B.
               Normalization and subtraction: two approaches to facilitate gene
               discovery
JOURNAL        Genome Res. 6 (9), 791-806 (1996)
MEDLINE        97044477
COMMENT        On May 7, 1998 this sequence version replaced gi:3119771.

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult 12-day-Embryo library. cDNA library preparation: M. Fatima
Bonaldo, Ph.D. Clone distribution: clones will be available through
Research Genetics

Seq primer: M13 Forward.
Location/Qualifiers
1. 429
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/notes="Vector: pUT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The U1-R-E1
library is a subcloned library derived from the U1-R-E0
library. The U1-R-E0 library consisted of a mixture of
individually tagged normalized libraries constructed from
8, 12 and 18-day embryo. The tag is a string of 3-5
nucleotides present between the NotI site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subcloned
library (U1-R-E1) was constructed as follows: PCR
amplified cDNA inserts from a pool of U1-R-E0 clones from
which 3' ESTs had been derived was used as a driver in a
hybridization with the U1-R-E0 library in the form of
single-stranded circles. The remaining single-stranded
circles (subcloned library) was purified by
hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the U1-R-E1
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, genome Research 6: 791-806,
1996)"
/db_xref="taxon:10116"
/map="18p11.1-q11.2"
/clone="U1-R-E1-rh-08-0-UI"
/clone_lib="U1-R-E1"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"

BASE COUNT    132 a      84 c      93 g      120 t
ORIGIN
Query Match   1.5%; Score 20; DB 26; Length 429;
Best Local Similarity 100.0%; Pred. No. 4.91e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 32 TATTATAACATATTAAAC 51
|||||
1348 TATTATAACATATTAAAC 1329

```



RESULT 4  
LOCUS AA956542 431 bp mRNA EST 07-FEB-1999  
DEFINITION UI-R-E1-fn-e-12-0-UI-st UI-R-E1 Rattus norvegicus cDNA clone  
UI-R-E1-fn-e-12-0-UI-3' similar to gi|29571591gb|AC004404|AC004404  
Mus musculus ma08m078, complete sequence [Mus musculus], mRNA  
sequence.  
ACCESSION AA956542  
NID 94239016  
VERSION AA956542.1 GI:4239016  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 431)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT On May 7, 1998 this sequence version replaced gi:3120237.  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
The sequence tag present in the cDNA between the NotI site and the  
oligo-dt track served to identify it as a clone from the normalized  
adult 12-day-embryo library. CDNA Library Preparation: M. Fatima  
Bonaldo, Ph.D. Clone distribution: clones will be available through  
Research Genetics  
Seq primer: M13 Forward.  
Location/Qualifiers  
1. 431  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/note="Vector: pRT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-E1  
library is a subtracted library derived from the UI-R-E0  
library. The UI-R-E0 library consisted of a mixture of  
individually tagged normalized libraries constructed from  
8, 12 and 18-day embryo. The tag is a string of 3-5  
nucleotides present between the NotI site and the  
oligo-dt track which allows identification of the library  
of origin of a clone within the mixture. The subtracted  
library (UI-R-E1) was constructed as follows: PCR  
amplified cDNA inserts from a pool of UI-R-E0 clones from  
which 3' ESTs had been derived was used as a driver in a  
hybridization with the UI-R-E0 library in the form of  
single-stranded circles. The remaining single-stranded  
circles (subtracted library) was purified by  
hydroxyapatite column chromatography, converted to  
double-stranded circles and electroporated into DH10B  
bacteria (Life Technologies) to generate the UI-R-E1  
library. This procedure has been previously described  
(Bonaldo, Lennon and Soares, genome Research 6: 791-806,  
1996)"  
/db\_xref="taxon:10116"  
/map="11g23"  
/clone="UI-R-E1-fn-e-12-0-UI"  
/clone\_lib="UI-R-E1"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
BASE COUNT 134 a 83 c 95 g 119 t  
ORIGIN  
Query Match 1.5%; Score 20; DB 26; Length 431;  
Best Local Similarity 100.0%; Pred. No. 4,91e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 39 TATTATAACATATTAAAC 58  
|||||  
Cp 1348 TATTATAACATATTAAAC 1329  
RESULT 5  
LOCUS AA943324 435 bp mRNA EST 16-JUN-1998  
DEFINITION EST198823 Normalized rat brain, Bento Soares Rattus sp. cDNA clone  
RRRA123 3' end, mRNA sequence.  
ACCESSION AA943324  
NID 93103240  
VERSION AA943324.1 GI:3103240  
KEYWORDS EST.  
SOURCE Rattus sp.  
ORGANISM Rattus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 435)  
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,  
Kerlavage,A.R. and Adams,M.D.  
TITLE Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat  
Gene Index  
JOURNAL Unpublished (1998)  
COMMENT On Apr 14, 1993 this sequence version replaced gi:716993.  
Contact: Lee, NH  
ATCC  
The Institute for Genomic Research  
9712, Medical Center Drive, Rockville, MD 20850, USA  
Tel: (301)-838-3529  
Fax: (301)-838-0208  
Email: nhlee@tigr.org  
Seq primer: M13-21  
Location/Qualifiers  
1. 435  
/organism="Rattus sp."  
/note="Organ: brain; Vector: pT73Pac; Site\_1: EcoRI;  
Site\_2: NotI"  
/db\_xref="ATCC (lnhost):2012840"  
/db\_xref="taxon:10118"  
/clone="RRRA123"  
/clone\_lib="Normalized rat brain, Bento Soares"  
BASE COUNT 134 a 85 c 102 g 114 t  
ORIGIN  
Query Match 1.5%; Score 20; DB 21; Length 435;  
Best Local Similarity 100.0%; Pred. No. 4,91e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 22 TATTATAACATATTAAAC 41  
|||||  
Cp 1348 TATTATAACATATTAAAC 1329  
RESULT 6  
LOCUS A1013962 460 bp mRNA EST 15-JUN-1998  
DEFINITION EST207517 Normalized rat spleen, Bento Soares Rattus sp. cDNA clone  
RSPBC29 3' end, mRNA sequence.  
ACCESSION A1013962  
NID 93226898  
VERSION A1013962.1 GI:3226898  
KEYWORDS EST.  
SOURCE Rattus sp.  
ORGANISM Rattus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 460)  
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,  
Kerlavage,A.R. and Adams,M.D.  
TITLE Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat  
Gene Index  
JOURNAL Unpublished (1998)



COMMENT On Jan 19, 1998 this sequence version replaced gi:2287332.

Contact: Lee, NH

ATCC

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@ligr.org

Seq primer: M13-21.

FEATURES  
source location/Qualifiers

1..460

/organism="Rattus sp."

/note="Organ: spleen; Vector: pT7T3Pac; Site\_1: EcoRI;

Site\_2: NotI

/db\_xref="taxon:10118"

/clone\_lib="RSPBC29"

/clone\_1lb="Normalized rat spleen, Bento Soares"

BASE COUNT 138 a 94 c 109 g 119 t

ORIGIN

Query Match 1.5%; Score 20; DB 21; Length 460;

Best Local Similarity 100.0%; Pred. No. 4.91e-06;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 15 TATTATTAACATATTAAAC 34

CP 1348 TATTATTAACATATTAAAC 1329

RESULT 7

LOCUS

DEFINITION EST220540 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone

ROVBY77 3' end, mRNA sequence.

ACCESSION A1176934

NID 93727572

VERSION A1176934.1 GI:3727572

KEYWORDS EST.

SOURCE Rattus sp.

ORGANISM Rattus sp.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 571)

Lee, N.H., Glodex, A., Chandra, I., Mason, T.M., Quackenbush, J.,

Kerlavage, A.R. and Adams, M.D.

Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat

Gene Index

Unpublished (1998)

On Jan 19, 1998 this sequence version replaced gi:2152166.

JOURNAL

COMMENT

Contact: Lee, NH

ATCC

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@ligr.org

Seq primer: M13-21.

FEATURES

source location/Qualifiers

1..571

/organism="Rattus sp."

/note="Organ: ovary; Vector: pT7T3Pac; Site\_1: EcoRI;

Site\_2: NotI

/db\_xref="taxon:10118"

/clone\_lib="Normalized rat ovary, Bento Soares"

/clone\_1lb="Normalized rat ovary, Bento Soares"

BASE COUNT 174 a 118 c 120 g 159 t

ORIGIN

Query Match 1.5%; Score 20; DB 23; Length 571;

Best Local Similarity 100.0%; Pred. No. 4.91e-06;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 15 TATTATTAACATATTAAAC 34

CP 1348 TATTATTAACATATTAAAC 1329

RESULT 8

LOCUS

DEFINITION EST203406 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone

REMG04 3' end, mRNA sequence.

ACCESSION A1008955

NID 94132969

VERSION A1008955.1 GI:4132969

KEYWORDS EST.

SOURCE Rattus sp.

ORGANISM Rattus sp.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 633)

Lee, N.H., Glodex, A., Chandra, I., Mason, T.M., Quackenbush, J.,

Kerlavage, A.R. and Adams, M.D.

Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat

Gene Index

Unpublished (1998)

On Jun 15, 1998 this sequence version replaced gi:3222787.

JOURNAL

COMMENT

Contact: Lee, NH

ATCC

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@ligr.org

Seq primer: M13-21.

FEATURES

source location/Qualifiers

1..633

/organism="Rattus sp."

/note="Vector: pT7T3Pac; Site\_1: EcoRI; Site\_2: NotI"

/db\_xref="taxon:10118"

/clone\_lib="REMG04"

/clone\_1lb="Normalized rat embryo, Bento Soares"

/dev\_stage="embryo 8, 12, 18 dpc"

BASE COUNT 190 a 139 c 126 g 178 t

ORIGIN

Query Match 1.5%; Score 20; DB 26; Length 633;

Best Local Similarity 100.0%; Pred. No. 4.91e-06;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 TATTATTAACATATTAAAC 41

CP 1348 TATTATTAACATATTAAAC 1329

RESULT 9

LOCUS

DEFINITION B68124 667 bp DNA GSS 17-JUN-1998

CT1978SK-A-802H9.7P CT1978SK Homo sapiens genomic clone A-802H9,

genomic survey sequence.

ACCESSION B68124

NID 92666834

VERSION B68124.1 GI:2666834

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

AUTHORS Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 667)

Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K.,

Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Kim, U.-J.,

Shizuya, H., Simon, M. and Venter, J.C.

Use of a human BAC End Sequence Database for Sequence-Ready Map

Building

Unpublished (1997)

JOURNAL

COMMENT







```

RESULT 12
LOCUS AQ233339 332 bp DNA GSS 25-SEP-1998
DEFINITION HS_2014_A2_G02_T7 CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2014 Col=4 Row=M, genomic survey
sequence.
ACCESSION AQ233339
NID 93650593
VERSION AQ233339.1 GI:3650593
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 332)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
TITLE Construction of a Characterized Clone Resource for Genomic
Sequencing
JOURNAL Unpublished (1998)
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2014 row: M column: 4
Class: BAC ends
High quality sequence stop: 332.
Location/Qualifiers
1. .332
/organism="Homo sapiens"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-COLI DH10B"
/db_xref="taxon:9606"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
BASE COUNT 98 a 57 c 55 g 119 t 3 others
ORIGIN
Query Match 1.4%; Score 19; DB 40; Length 332;
Best Local Similarity 100.0%; Pred. No. 2.10e-04;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 34 TTATTCAAGATATTAT 52
Oy 552 TTATTCAAGATATTAT 570

RESULT 13
LOCUS AA080511 376 bp mRNA EST 15-FEB-1997
DEFINITION m8861.11 StrataGene mouse Tcell 937311 Mus musculus cDNA clone
IMAGE:551180 5', mRNA sequence.
ACCESSION AA080511
NID 91619483
VERSION AA080511.1 GI:1619483
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 376)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,T., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The Mashu-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1393265.

FEATURES
SOURCE
Contact: Marra M/Mouse EST Project
Mashu-HMI Mouse EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800.
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:331972
Seq primer: -28n13 rev1 EP from Amersham
High quality sequence stop: 346.
Location/Qualifiers
1. .376
/organism="Mus musculus"
/note="Organ: blood; Vector: pBluescript SK-; Site:1;
EcORI: Site:2; XhoI: Cloned unidirectionally. Primer:
Oligo dT. M30 CD4+ cells. Average insert size: 1.0 kb;
Uni-ZAP XR Vector: ~5' adaptor sequence: 5' GAATTCGACGACG
3' ~3' adaptor sequence: 5' CTCGAGTGTGTGTGTGTGTGT 3'"
/db_xref="taxon:10090"
/clone_lib="Stratagene mouse Tcell 937311"
/clone_type="Tcell"
/dev_stage="M30 CD4+ cells"
/lab_host="SOLR (kanamycin resistant)"
BASE COUNT 109 a 60 c 67 g 139 t 1 others
ORIGIN
Query Match 1.4%; Score 19; DB 36; Length 376;
Best Local Similarity 100.0%; Pred. No. 2.10e-04;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 226 CATTATGCTCAATATT 244
Oy 705 CATTATGCTCAATATT 723

RESULT 14
LOCUS AQ115573 383 bp DNA GSS 28-AUG-1998
DEFINITION R0C111-57N20.TK R0C111 Homo sapiens genomic clone R-57N20, genomic
survey sequence.
ACCESSION AQ115573
NID 93491694
VERSION AQ115573.1 GI:3491694
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 383)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Sub,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
For clone availability, please contact Pieter de Jong
(pieter@tigr.org, med.buffalo.edu). Clones may be purchased from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.
Class: BAC ends.
Location/Qualifiers
1. .383
/organism="Homo sapiens"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;

```



RPC11 Human Male BAC Library"

/db\_xref="taxon:9606"

/clone="R-57N20"

/clone\_lib="RPC11"

/sex="Male"

/cell\_type="lymphocytes"

BASE COUNT 142 a 48 c 93 g 100 t

ORIGIN

Query Match 1.4%; Score 19; DB 39; Length 383;

Best Local Similarity 100.0%; Pred. No. 2.10e-04; Mismatches 0; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 72 GGACATGAGTCAAGAT 90  
 |||||  
 OY 94 GGACATGAGTCAAGAT 112

RESULT 15  
 LOCUS AQ093108 399 bp DNA GSS 26-AUG-1998

DEFINITION HS\_3020\_B1\_A12\_T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3020 COL-23 Row-B, genomic survey

sequence

ACCESSION AQ093108

NID 93464555

VERSION AQ093108.1 GI:3464555

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 399)

Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,

Tralcoff,R., Abajlan,C., Blanchard,A., West,A. and Hood,L.E.

Construction of a Characterized Clone Resource for Genomic

Sequencing: Generation and Preliminary Analysis of 20,000 Sequence

Tagged Connectors

Unpublished (1997)

CONTACT: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 3020 Row: B Column: 23

Class: BAC ends

High quality sequence stop: 399.

Location/Qualifiers

1. .399

/organism="Homo sapiens"

/note="Organ: Sperm; Vector: pBelobAC11; BAC Clones in

E-COL1 DH10B"

/db\_xref="taxon:9606"

/clone="plate-3020 COL-23 Row-B"

/clone\_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

BASE COUNT 166 a 70 c 61 g 100 t 2 others

ORIGIN

Query Match 1.4%; Score 19; DB 39; Length 399;

Best Local Similarity 100.0%; Pred. No. 2.10e-04;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 294 CCTGCACCATTAATAA 312

|||||

Cp 1274 CCTGCACCATTAATAA 1256

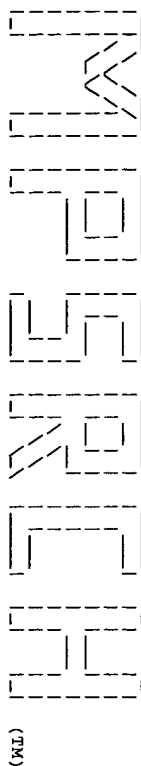
Search completed: Sat Nov 27 14:08:54 1999

Job time : 2299 secs.



**THIS PAGE BLANK (USPTO)**





(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mparch\_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 14:17:20 1999; MasPar time 1211.94 Seconds

Tabular output not generated. 1509.527 Million cell updates/sec

Title: >US-09-103-287-3  
Description: (1-660) from US09103287.seq  
Perfect Score: 660  
N.A. Sequence: 1 ATTAAAGATTCGATGACA.....GCGTTTAAATGTTTATTA 660  
Comp: TAAATTCTAAGCCTACTGT.....CGCAAAATTATACAAATATT

Scoring table: TABLE jmetric

Gap 60

Mmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 138595363 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

Database:

emb158  
1:em\_ba1 2:em\_ba2 3:em\_fun 4:em\_hlg 5:em\_hum1 6:em\_hum2  
7:em\_in 8:em\_om 9:em\_or 10:em\_ov 11:em\_pat 12:em\_ph  
13:em\_pl 14:em\_ro 15:em\_sts 16:em\_vi  
genbank111  
17:gb\_ba1 18:gb\_ba2 19:gb\_hlg1 20:gb\_hlg2 21:gb\_in1  
22:gb\_in2 23:gb\_om 24:gb\_ov 25:gb\_pat 26:gb\_ph 27:gb\_pl1  
28:gb\_pl2 29:gb\_pr1 30:gb\_pr2 31:gb\_pr3 32:gb\_ro  
33:gb\_st 34:gb\_sts 35:gb\_sy 36:gb\_un 37:gb\_vi

Statistics: Mean 8.797; Variance 3.217; scale 2.734

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %	Match Length	DB ID	Description	Pred. No.
1	494	74.8	1314	17	Staphylococcus aureus	0.00e+00
2	21	3.2	31281	21	Caenorhabditis elegans	2.86e-01
3	20	3.0	4043	17	BACSKW01	1.47e+00
4	20	3.0	12188	18	Haemophilus influenzae	1.47e+00
5	20	3.0	177411	31	Ac005696	1.47e+00
6	19	2.9	351	34	human STS CHIC.GCT3806	7.20e+00
7	19	2.9	524	34	human STS CHIC.GCT3806	7.20e+00
8	19	2.9	2057	17	MC097	7.20e+00
9	19	2.9	2650	29	HUMIGV3BA	7.20e+00
10	19	2.9	14459	22	AE001383	7.20e+00
11	19	2.9	16826	23	MTPVNA	7.20e+00
12	19	2.9	36539	21	CEK08H2	7.20e+00
13	19	2.9	39030	21	CE41D3	7.20e+00

14	19	2.9	41275	21	CELF47D12	Caenorhabditis elegans	7.20e+00
15	19	2.9	70475	27	AB018112	Arabidopsis thaliana g	7.20e+00
16	19	2.9	72839	28	ATAC002409	Arabidopsis thaliana c	7.20e+00
17	19	2.9	74192	20	AC005504	Plasmodium falciparum	7.20e+00
18	19	2.9	76618	21	AC004657	Drosophila melanogaster	7.20e+00
19	19	2.9	110000	19	CEV43D4	Caenorhabditis elegans	7.20e+00
20	19	2.9	148020	19	HS796E4	Human DNA sequence ***	7.20e+00
21	19	2.9	151052	20	AC005506	Plasmodium falciparum	7.20e+00
22	19	2.9	175066	31	AC005900	Homo sapiens chromosome	7.20e+00
23	19	2.9	268531	19	CEY51H4	Caenorhabditis elegans	7.20e+00
24	19	2.9	297070	19	CEY47H10	Caenorhabditis elegans	7.20e+00
25	19	2.9	297070	19	CEY47H10	Caenorhabditis elegans	7.20e+00
26	18	2.7	23360	21	DHMHHC	D. hydei MMHC gene.	3.32e+01
27	18	2.7	30147	22	DMU66884	Drosophila melanogaster	3.32e+01
28	18	2.7	31800	21	CEY21B10	Caenorhabditis elegans	3.32e+01
29	18	2.7	38070	21	CEY55H12	Caenorhabditis elegans	3.32e+01
30	18	2.7	40050	22	CELC08G5	Caenorhabditis elegans	3.32e+01
31	18	2.7	75787	4	AC005426	*** SEQUENCING IN PROG	3.32e+01
32	18	2.7	102258	29	HS295C6	Human DNA sequence firo	3.32e+01
33	18	2.7	133060	20	AC006474	Homo sapiens clone DJ0	3.32e+01
34	18	2.7	133195	28	AC007123	Arabidopsis thaliana B	3.32e+01
35	18	2.7	135182	22	AC007121	Drosophila melanogaster	3.32e+01
36	18	2.7	144158	19	HS179E13	Human DNA sequence ***	3.32e+01
37	18	2.7	163418	20	AC007026	Homo sapiens clone DJ0	3.32e+01
38	18	2.7	187755	19	AC006718	Caenorhabditis elegans	3.32e+01
39	18	2.7	207139	19	CEY17G7	Caenorhabditis elegans	3.32e+01
40	18	2.7	210137	31	AC006504	Homo sapiens chromosome	3.32e+01
41	18	2.7	213541	31	HUAC004381	Homo sapiens Chromosome	3.32e+01
42	18	2.7	312430	17	RPKX02	Rickettsia prowazekii	3.32e+01
43	18	2.7	319000	17	AP000006	Pyrococcus horikoshii	3.32e+01
44	18	2.7	337565	19	CEY47D3	Caenorhabditis elegans	3.32e+01
45	18	2.7	342347	19	CEY39B6	Caenorhabditis elegans	3.32e+01

## ALIGNMENTS

RESULT 1	AF034076	1314 bp	DNA	BCT	26-NOV-1997
LOCUS	Staphylococcus aureus	UDP-N-acetylglucosyl-L-alanine synthetase			
DEFINITION	(murC) gene, complete cds.				
ACCESSION	AF034076				
VERSION	92642658				
KEYWORDS	AF034076.1	GT-2642658			
SOURCE	Staphylococcus aureus.				
ORGANISM	Bacteria; Firmicutes; Low G+C gram-positive bacteria;				
REFERENCE	Bacillaceae; Staphylococcus.				
AUTHORS	1 (bases 1 to 1314)				
TITLE	Lowe,A.M. and Deresiewicz,R.L.				
JOURNAL	Cloning and sequencing of Staphylococcus aureus murC, a gene				
AUTHORS	involved in cell wall biosynthesis				
TITLE	unpublished				
JOURNAL	2 (bases 1 to 1314)				
AUTHORS	Lowe,A.M. and Deresiewicz,R.L.				
TITLE	Submitted (11-NOV-1997) Channing Laboratory, Brigham and Women's				
JOURNAL	Hospital and Harvard Medical School, 181 Longwood Ave, Boston, MA				
FEATURES	02115, USA				
source	Location/Qualifiers				
gene	1..1314				
CDS	/organism="Staphylococcus aureus"				
	/db_xref="taxon:1280"				
	1..1314				
	/gene="murC"				
	1..1314				
	/gene="murC"				
	/function="cell wall biosynthesis"				
	/note="MurC: UDP-N-acetylglucosyl-L-alanine ligase"				
	/codon_start=1				
	/transl_table=11				
	/product="UDP-N-acetylglucosyl-L-alanine synthetase"				
	/protein_id="AAB87090.1"				







gene: cDNA EST yk353f12.3 comes from this gene; cDNA EST yk353f12.5 comes from this gene; cDNA EST yk310a7.3 comes from this gene; cDNA EST yk310a7.5 comes from this gene; cDNA EST yk280d11.3 comes from this gene; cDNA EST yk280d11.5 comes from this gene; cDNA EST yk296a2.3 comes from this gene"

gene  
/codon\_start=1  
/protein\_id="CA92434.1"  
/db\_xref="PID:e1343689"  
/db\_xref="PID:g3873814"  
/db\_xref="GI:3873814"  
/translation="MAATGIEITFAIRNARATPPSADGSLKIFSLDKMEVEKAEET  
EATYQSDAFETKELNANRIKMOYRTIVLNSSTKINGSGTADFDISMFAE  
GGRILLILASPLAPCSLEIVGVDPYVSOFLRIRFSGIASHSPDVLPLQLYILSIP  
TEERTIMLYKSEMRETKYEPWAASIPLEFLNYESIOLHYVNYPPHDDOLVGH  
CTTTLTLOOGVGHNSVYMLMEPNKRIKKEICIELKSELNGSPPEPMENNNKIO  
LTSALDITRNSGNPNQSSLAHITHRQPSPIYELALQYTPPLATILPQNPHICALG  
FEAKVQPGGALQISHCFCNLGTPDPRVREGGLISATRTAVMGLPAPDPDESEVI  
YFMSKFAKASERHVLGYFLVLIYSDGPNALNNKRSIDAIVDASPMPMSIIGVGM  
QDRHSPMRLKELTLKHSGRLLVQNYSYVPSDLESSDVLAIMPQMAQMKRMFH  
FDPK"

gene  
/gene="C01F6.2"  
/join(4953..5032,5081..5575,5663..5810,5860..5985,  
6033..6264,6332..6585)  
/gene="C01F6.2"  
/codon\_start=1  
/protein\_id="CA92435.1"  
/db\_xref="PID:e1343690"  
/db\_xref="PID:g3873815"  
/db\_xref="GI:3873815"  
/db\_xref="GI:3873815"  
/db\_xref="SPTREMBL:Q17560"  
/translation="MEHFADPHIVYERSEERPMYSWGIILAVILYVMSLYTLEA  
GYRISIRKIMLSMLDPOOSTVSTPRCLSPNNSKSSPOTKTPDVIROKYPNMEP  
NCVTRPVYPTKTIPECTAVPRLNSEDILLRDCIGCKKINGSLEKDKKVESE  
KEKIKETKSSPQMTLDEVOQDASSNENIRPTMAPSPVNHSTQPSNRKSTGSDS  
NLANAPNSHSCDOINERVLQIAGCNLDEDRTDAGACDAISPDGSGMDHJGSDS  
FEGKIODIOLQSDNSILEATSDENLSMSWTSLSLAIEPSFETDVQASE  
TSHCEPNLGAQOTVNEODADIVIDSHLPAFSENEGSGSLVKGGOOKVAFRMTDE  
KRTYVDSVTLNCSFECNMNIPMKRNEKMEVCIEELPGMDIYLYNTRP"

gene  
/complement(6798..8876)  
/gene="C01F6.3"  
/complement(join(6798..6884,7346..7969,8083..8202,  
8250..8346,8394..8428,8525..8737,8784..8876))  
/gene="C01F6.3"  
/note="similar to cytochrome P450; cDNA EST EMBL:C13310  
comes from this gene; cDNA EST EMBL:C11235 comes from this  
gene"

gene  
/codon\_start=1  
/protein\_id="CA92436.1"  
/db\_xref="PID:e1343691"  
/db\_xref="PID:g3873816"  
/db\_xref="GI:3873816"  
/db\_xref="SPTREMBL:Q27466"  
/translation="MGVILAVLASATVIANLLYKHLRMOALKHLNPRSYPITIG  
GLITRDPGCFMNOVIGMGYLPDPMPCLMTGSPCLMYSGDVEAFVSTKILNR  
ILVOKLCCGADERDVLSIALCTLIDICETMSGALAEENNEYAWAHVHTIKL  
ISKRNPLCMANSEIYNLXGFIINKENDYKMGRLAPFDLLLEWNSQOMETVQV  
ECNITMLBGHDTSTGLMAVHLGHNHPVOKRVOAELEDEVGDDDEDVTELSRKY  
LECALKEALRLPESVALITRELSDQVIGFNIPIKGVFLLLYLVHRPPAKMOPDV  
FDPDRFHPENSIGRSFAIFPSAGSRNIGORFALMEKVIAMHLRNFNIKAVELM  
HEVRPMELIIVRPVPIHMKLTRRPIASP"

gene  
/gene="fem-3"  
/join(10227..10297,10396..10740,10786..10864,11236..11583,  
11666..11756,12137..12369)  
/gene="fem-3"  
/note="identity with sex-determining protein FEM-3  
(SM:FEM3 CAELI)"  
/codon\_start=1  
/protein\_id="CA92437.1"  
/db\_xref="PID:e1343692"

gene  
/db\_xref="PID:g3873817"  
/db\_xref="GI:3873817"  
/db\_xref="SWISS-PROT:P34691"  
/translation="MEVPGSDVDYADDEFRAOKLKLKNVYKFRAMRPFDEYGVN  
LTVDDLMLISGIFLORGLTGATGYDDSLDPMPEFSIDRELETISLITAGA  
GYERDDETRFVFORMTKTYTCNFMDLARRNCKVOITREFLODLIHRDBFRYV  
FRQPLNDPNCOTAFSNHTLSLHLIFNIGMPQFMNSQNTNATIRGRFLVONIA  
AFYPPETWNDESKYHTFEVPRGTSEKYYARFHALMNPENLITVLMWLAK  
LCILEIYVHTIMCDITGFGGLPRIEHRFLAMENEDIFDLAIDFISIRLODISP  
FEISRSPLDVSGYEYTIKKRKIDIEYQNRFEVAYSDVIRIMNYATDCSRK"  
13971..14691  
/gene="C01F6.5"  
/join(13971..14172,14222..14691)  
/gene="C01F6.5"  
/note="similar to RNA recognition motif. (aka RRM, RBD, or  
RNP domain); cDNA EST EMBL:U00641 comes from this gene;  
cDNA EST EMBL:U00642 comes from this gene; cDNA EST  
EMBL:D72572 comes from this gene; cDNA EST EMBL:D7485  
comes from this gene; cDNA EST yk269c10.5 comes from this  
gene; cDNA EST yk269c10.5 comes from this gene"

gene  
/codon\_start=1  
/protein\_id="CA92438.1"  
/db\_xref="PID:e1343693"  
/db\_xref="PID:g3873818"  
/db\_xref="GI:3873818"  
/db\_xref="SPTREMBL:Q17561"  
/translation="MARAIADAPLASGIRRTKSPIKRASVGVKRRFSTGAPSR  
RPTAQRROSGGIRTVDRISIPANERVINLSNAHVHSGDLRLQFAEFKIRVSV  
NPNHGKPVGCGDVTLPRKHADRIQKAGVSLCKETKFAITDSANANVPPEAP  
RRAPSNGRPQARPRQNVKSGKPTGGAGQAKKLAKKPRKADQPKTLELDALDY  
YMSKA"

gene  
/gene="C01F6.6"  
/join(24126..24194,24453..24532,24780..25095)  
/gene="C01F6.6"  
/note="similar to PDZ domain (also known as DHR or GLGF).;  
cDNA EST EMBL:U14584 comes from this gene"

gene  
/codon\_start=1  
/protein\_id="CA92439.1"  
/db\_xref="PID:e1343694"  
/db\_xref="PID:g3873819"

...  
Note: remainder of annotations omitted.

Query Match 3.2%; Score 21; Length 31281;  
Best Local Similarity 100.0%; Pred. No. 2,86e-01;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1108 TTTTAAATGAATTGCAGAA 1128  
OY 387 TTTTAAATGAATTGCAGAA 407  
|||||  
|||||

RESULT 3  
LOCUS BACSKW01 4043 bp DNA  
DEFINITION Bacillus thuringiensis DNA.  
ACCESSION D86064  
NID 91389567  
VERSION D86064.1 GI:1389567  
KEYWORDS  
SOURCE Bacillus thuringiensis (sub-species:sotto, strain:SKW01-10-2-06)  
DNA.  
ORGANISM Bacillus thuringiensis  
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;  
Bacillus.  
REFERENCE 1 (bases 1 to 4043)  
AUTHORS Sasaki,J.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUN-1996) to the DBJ/EMBL/GenBank databases. Jun  
Sasaki, Hokkaido University, Applied life science; N9, W9, Sapporo,  
Hokkaido 060, Japan (E-mail:JSS@el.hines.hokudai.ac.jp,  
Tel.81-11-706-2402)  
2 (bases 1 to 4043)  
REFERENCE  
AUTHORS Sasaki,J.



TITLE Bacillus thuringiensis serovar sotto SKM01-10-2-06  
JOURNAL Unpublished (1996)  
FEATURES Location/Qualifiers  
source 1..4043  
/organism="Bacillus thuringiensis"  
/strain="SKM01-10-2-06"  
/sub\_species="sotto"  
/db\_xref="taxon:1428"  
BASE COUNT 1430 a 578 c 751 g 1284 t  
ORIGIN  
Query Match 3.0%; Score 20; DB 17; Length 4043;  
Best Local Similarity 100.0%; Pred. No. 1,47e+00;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1868 ATACAAGATTATGATGA 1887  
|||||  
Oy 482 ATACAAGATTATGATGA 501  
RESULT 4  
LOCUS U32809 12188 bp DNA BCT 09-SEP-1998  
DEFINITION Haemophilus influenzae Rd section 124 of 163 of the complete  
genome  
ACCESSION U32809.L42023  
NID U3212217  
VERSION U32809.1 GI:3212217  
KEYWORDS  
SOURCE  
ORGANISM Haemophilus influenzae Rd.  
Haemophilus influenzae Rd.  
Eubacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus.  
REFERENCE  
AUTHORS 1 (bases 1 to 12188)  
Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A.,  
Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J., Dougherty, B.A.,  
Merrick, J.M., McKenney, K., Sutton, G.G., FitzHugh, W., Fields, C.A.,  
Gocayne, J.D., Scott, J.D., Shiley, R., Liu, L.I., Glodek, A.,  
Kellley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E.,  
Cotton, M.D., Utterback, T., Hanna, M.C., Nguyen, D.T., Saudek, D.M.,  
Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L.,  
Geoghegan, N.S., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M.,  
Smith, H.O. and Venter, J.C.  
TITLE Whole-genome random sequencing and assembly of Haemophilus  
influenzae Rd  
JOURNAL Science 269 (5223), 496-512 (1995)  
MEDLINE 95350630  
REFERENCE  
AUTHORS 2 (bases 1 to 12188)  
Tatusov, R.L., Mushegian, A.R., Bork, P., Brown, N.P., Hayes, W.S.,  
Borodovsky, M., Rudd, K.E. and Koonin, E.V.  
TITLE Metabolism and evolution of Haemophilus influenzae deduced from a  
whole-genome comparison with Escherichia coli  
JOURNAL Curr. Biol. 6 (3), 279-291 (1996)  
MEDLINE 96398784  
REFERENCE  
AUTHORS 3 (bases 1 to 12188)  
White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.  
TITLE Direct Submission  
JOURNAL Submitted (25-JUL-1995) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
REFERENCE  
AUTHORS 4 (bases 1 to 12188)  
White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.  
TITLE Direct Submission  
JOURNAL Submitted (27-SEP-1997) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
REMARK The H. influenzae sequence has been updated by R. Fleischmann. New  
database matches have been assigned, product names have been  
improved, and a number of frame shifts have been corrected. We  
gratefully acknowledge the work of Tatusov et. al. We have  
incorporated their annotation into the /notes fields of the  
corresponding H. influenzae genes  
REFERENCE  
AUTHORS 5 (bases 1 to 12188)  
White, O., Clayton, R.A., Kerlavage, A.R., Fleischmann, R.D.,  
Peterson, J., Hickey, E., Dodson, R. and Winn, M.  
TITLE Direct Submission

JOURNAL Submitted: (28-MAY-1998) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
REMARK The whole genome was shifted by 588 nucleotides for a new start  
COMMENT On Jun 12, 1998 this sequence version replaced gi:1574746.  
FEATURES Location/Qualifiers  
source 1..12188  
/organism="Haemophilus influenzae Rd"  
/db\_xref="taxon:71421"  
gene 288..674  
/gene="Hii288"  
CDS 288..674  
/gene="Hii288"  
/note="similar to SP:P09170 GB:X13270 PID:42218 PID:42222  
PID:606107 percent identity: 67.44; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="ribosome binding factor A (rfa)"  
/protein\_id="AAC22937.1"  
/db\_xref="PID:g1574747"  
/db\_xref="GI:1574747"  
/translation="MAREFRSDRVAQEIQKETAIIQREVKDPRIGMTVSDVEVS  
DISAKIEVTFEPDDEMATQCGKGLKSPYRSLGKAMRLRIYPIRIFIDSL  
VEGRMSNLVTNVYREDEKKEESN"  
674..1612  
/gene="Hii289"  
674..1612  
/gene="Hii289"  
/note="similar to SP:P09171 GB:X13270 PID:42219 PID:42223  
PID:606106 percent identity: 58.44; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="tRNA pseudouridine 55 synthase (trub)"  
/protein\_id="AAC22938.1"  
/db\_xref="PID:g1574748"  
/db\_xref="GI:1574748"  
/translation="MSRPRKRWPDVGVFLDLKPOGSSNDIMQKVRLEFOANKAGHT  
GALDPLATGMPLCEGATFESGILDAKRYLTAKIGRTSDSGGVETREYN  
LENOQLITALEPFGDILQVTPMSALKNHKGPIYEVARGITVEREARRITTELE  
IENAPRLTEVHCSTKTYRTLYVDYSGVLCGCAHYTMARRVADYPAEMQPIRE  
LQIAESFPISEIDRLPLPTDAVSKPALHLDQSKALGFQGRKAFANQDLSGV  
RLPSAENLFTGVNRRREYSPPTINVTIRITSLPFL"  
1672..2805  
/gene="Hii290"  
1672..2805  
/gene="Hii290"  
/note="similar to SP:Q02287 GB:M74135 GB:X60420 PID:415010  
PID:43345 percent identity: 58.86; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="chorismate mutase / prephenate dehydrogenase  
(tyrA)"  
/protein\_id="AAC22939.1"  
/db\_xref="PID:g1574749"  
/db\_xref="GI:1574749"  
/translation="MSFEMALKDURSEIDSLRELIDLFKRLIEVSGVKVHROHGL  
PIYAPEREIMQARRLEARKAGISADLIEDVYRRREXYANQNGEFTNIDHK  
IIVYGCGKLGGLFARILRASGTPISILDEEDNAVAESTIANDVYVSPINLTLET  
IERKPYLTENMLADIQTSVREPLAMLEVHTGAVGLHPMGADIASAKOVVYVC  
DGRPERYEWLLEIDQITGAKIYQNTNTEHDHMYTIOALRHFSTFANGJLHLSKOPIN  
LANLALASPIYRLELMIGRILPAQDAELVADIIMDKSEMLAVIETLKYTDALTFE  
ENNDROGFIAPFKVRMPEDYSQFLKESROLQANDLKQC"  
complement(2838..3677)  
/gene="Hii291"  
complement(2838..3677)  
/gene="Hii291"  
/note="similar to PID:882689 GB:U00096 SP:O46920  
PID:1789158 percent identity: 61.59; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1







```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hRPK_74_E_22"
/clone_lib="RPCI-11 human BAC library"
/map="17"
/chromosome="17"
repeat_region complement(1..60)
repeat_region /rpt_family="THE1A"
repeat_region complement(64..172)
repeat_region /rpt_family="FLAM_C"
repeat_region 332..368
repeat_region /rpt_family="AT_rich"
repeat_region 745..846
repeat_region /rpt_family="Charlie4"
repeat_region 974..1288
repeat_region /rpt_family="AluJo"
repeat_region 1299..1595
repeat_region /rpt_family="AluSx"
repeat_region complement(1900..2178)
repeat_region /rpt_family="AluSg"
repeat_region complement(2181..6183)
repeat_region /rpt_family="L1PBa"
repeat_region 6288..6599
repeat_region /rpt_family="AluSx"
repeat_region complement(6603..7203)
repeat_region /rpt_family="L1PBa"
repeat_region complement(7231..7423)
repeat_region /rpt_family="L1MC5"
repeat_region complement(7424..7556)
repeat_region /rpt_family="FLAM_C"
repeat_region complement(7557..7893)
repeat_region /rpt_family="L1MC5"
repeat_region complement(7895..8211)
repeat_region /rpt_family="MER2"
repeat_region complement(8221..8332)
repeat_region /rpt_family="L1MC5"
repeat_region complement(8487..8795)
repeat_region /rpt_family="AluSc"
repeat_region complement(8751..8803)
repeat_region /note="Single-stranded coverage."
repeat_region 8915..9461
repeat_region /rpt_family="L2"
repeat_region complement(9462..9763)
repeat_region /rpt_family="AluJb8"
repeat_region 9764..9965
repeat_region /rpt_family="L2"
repeat_region 10065..10208
repeat_region /rpt_family="L2"
repeat_region 10209..10503
repeat_region /rpt_family="AluSc"
repeat_region 10504..10598
repeat_region /rpt_family="L2"
repeat_region complement(10683..10982)
repeat_region /rpt_family="AluJo"
repeat_region 11146..11459
repeat_region /rpt_family="AluJb"
repeat_region complement(11774..12025)
repeat_region /rpt_family="AluSx"
repeat_region complement(12059..12351)
repeat_region /rpt_family="AluJo"
repeat_region 12371..12453
repeat_region /rpt_family="L2"
repeat_region 12659..12918
repeat_region /rpt_family="AluJo"
repeat_region complement(13821..14133)
repeat_region /rpt_family="AluSg"
repeat_region complement(14224..14523)
repeat_region /rpt_family="AluY"
repeat_region complement(14915..15206)
repeat_region /rpt_family="AluSg"
repeat_region 15214..15242
repeat_region /rpt_family="AT_rich"
repeat_region complement(15258..15777)
```

```
repeat_region /rpt_family="L1MC4"
repeat_region complement(15778..16043)
repeat_region /rpt_family="AluSg"
repeat_region complement(16044..16460)
repeat_region /rpt_family="L1MC4"
repeat_region complement(16461..16783)
repeat_region /rpt_family="AluSx"
repeat_region complement(16784..17170)
repeat_region /rpt_family="L1MC4"
repeat_region complement(17208..17349)
repeat_region /rpt_family="MER20"
repeat_region complement(17403..17565)
repeat_region /rpt_family="L1MC4"
repeat_region 17566..17942
repeat_region /rpt_family="MER46B"
repeat_region complement(17943..17994)
repeat_region /rpt_family="L1MC4"
repeat_region 18201..18222
repeat_region /rpt_family="AT_rich"
repeat_region complement(18406..18743)
repeat_region /rpt_family="AluJo"
repeat_region complement(18744..18879)
repeat_region /rpt_family="AluJb"
repeat_region complement(18880..19184)
repeat_region /rpt_family="AluSg"
repeat_region complement(19195..19478)
repeat_region /rpt_family="AluJo"
repeat_region 19585..19891
repeat_region /rpt_family="AluSg"
repeat_region 20232..20289
repeat_region /rpt_family="L1P"
repeat_region 20280..20585
repeat_region /rpt_family="AluJb"
repeat_region 20586..20640
repeat_region /rpt_family="L1P"
repeat_region 20641..20921
repeat_region /rpt_family="AluSg"
repeat_region 20922..20949
repeat_region /rpt_family="(TA)n"
repeat_region complement(20950..21228)
repeat_region /rpt_family="AluSg1"
repeat_region 21229..21571
repeat_region /rpt_family="L1P"
repeat_region 21572..21750
repeat_region /rpt_family="L1M4"
repeat_region 21751..21996
repeat_region /rpt_family="AluSx"
repeat_region 21829..21851
repeat_region /note="Single-stranded coverage."
repeat_region 21997..22274
repeat_region /rpt_family="L1M4"
repeat_region 22275..22392
repeat_region /rpt_family="AluSg"
repeat_region 22393..22709
repeat_region /rpt_family="AluSx"
repeat_region 22710..22880
repeat_region /rpt_family="AluSg"
repeat_region 22881..22976
repeat_region /rpt_family="L1M4"
repeat_region complement(22977..23047)
repeat_region /rpt_family="THE1C"
repeat_region 23048..23327
repeat_region /rpt_family="AluY"
repeat_region 23328..23354
repeat_region /rpt_family="(TAA)n"
repeat_region complement(23362..23665)
repeat_region /rpt_family="THE1C"
repeat_region 23666..23715
repeat_region /rpt_family="L1M4"
```

...
Note: remainder of annotations omitted.

Query Match 3.0%; Score 20; DB 31; Length 177411;



Best Local Similarity 100.0%; Pred. No. 1.47e+00;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 123270 TGAAGATCTATTAATGTAT 123289  
|||||  
OY 526 TGAAGATCTATTAATGTAT 545

RESULT 6 351 bp DNA STS 23-MAR-1999  
LOCUS 647692  
DEFINITION 225793.1 zebrafish AB Danio rerio STS genomic clone 225793 5',  
sequence tagged site.

ACCESSION 647692  
NID 94494300  
VERSION 647692.1 GI:4494300  
KEYWORDS STS.  
SOURCE zebrafish.

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
Cyprinidae; Cyprininae; Rasbora; Danio.

REFERENCE 1 (bases 1 to 351)  
Ziniti,J., Jackson,D.G. and Fishman,M.C.  
TITLE A zebrafish microsatellite map  
JOURNAL Unpublished (1999)

COMMENT

Contact: Mark C. Fishman  
Cardiovascular Research Center  
Massachusetts General Hospital  
Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA  
Fax: 617/7265806

Email: fishman@mgh.cvr.harvard.edu  
http://zebrafish.mgh.harvard.edu  
Primer A: TCTGACCTTAGCAGATGA  
Primer B: AAGTCCTCTGCGGATCTT  
STS size: 229

PCR Profile:

Presoak: 94 degrees C for 5.0 minutes  
Denaturation: 94 degrees C for 1.0 minute  
Annealing: 58 degrees C for 1.0 minute  
Polymerization: 72 degrees C for 1.5 minute  
PCR Cycles: 27  
Thermal Cycler: MJ Research PTC-100

Protocol:

Template: 10 ng  
Primer: each 375 nM  
dNTPs: each 200 uM  
Tag Polymerase: 0.034 units/uL  
Total Vol: 10 uL

Buffer:

MgCl2: 1.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 8.3.

Location/Qualifiers

1..351

/organism="Danio rerio"

/strain="AB"

/note="Vector: m13mp19 with added BstXI site; V-type:  
Phage; Genomic DNA from a single adult zebrafish of AB  
strain was digested with AluI, CaeIII, HaeIII, NlaIV, or  
RsaI. Fragments in the range of 250-500 bp were gel  
purified and a BstXI linker was added. The fragments were  
cloned into a modified M13mp19 vector and transformed  
into E. coli DH5alpha. Microsatellite sequences were  
screened with labeled d(CA)15 and d(CT)15 oligonucleotide  
probes."

/db\_xref="taxon:7955"  
/clone="225793"  
/clone\_1bp="zebrafish AB"  
/sex="F"  
/dev\_stage="Adult"

STS /lab\_host="DH5alphaF'10"  
primer\_bind 8..236  
primer\_bind 3..27  
BASE COUNT complement(217..236)  
98 a 94 c 80 g 79 t

Query Match 2.9%; Score 19; DB 34; Length 351;  
Best Local Similarity 100.0%; Pred. No. 7.20e+00;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 292 CACATACATCAACACA 310  
|||||  
Cp 76 CACATACATCAACACA 58

RESULT 7

LOCUS 609531 524 bp DNA STS 14-AUG-1995  
DEFINITION human STS CHLC.GCT3B06.P11066 clone GCT3B06.

ACCESSION 609531  
NID 9941380  
VERSION 609531.1 GI:941380

KEYWORDS STS sequence; primer; sequence tagged site.  
SOURCE human vector-pvCP1 host-E.coli dut-ung+ (DH10B) Marker Selected  
genomic DNA prepared from XY individual of French nationality.  
Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Eukaryotes; Metazoa; Chordata;  
Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Channata;  
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;  
Cathartini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 524)  
Murray,J., Sheffield,V., Weber,J.L., Duyk,G. and Buetow,K.H.  
TITLE Cooperative Human Linkage Center  
JOURNAL Unpublished (1995)  
Synonyms: GCT3B06, CHLC.GCT3B06.T11065  
COMMENT Contact: Dr. Jeffrey C. Murray

Doi

The University of Iowa  
Department of Pediatrics, Iowa City, IA 52242, USA  
Tel: (319) 356-3508  
Fax: (319) 356-3347  
Email: jeff-murray@uiowa.edu

Primer A: ATTGACTGAGATGATCG  
Primer B: TCCCTAGGCTTGGCTTCT  
STS size: 141  
PCR Profile:

denature: 30 seconds at 94 degrees C  
annealing: 75 seconds at 55 degrees C  
extension: 15 seconds at 72 degrees C  
PCR cycles: 27  
extension: 6 minutes at 72 degrees C

Protocol:

Template: 30ng genomic DNA  
Primer: each 1.5 pmole  
dNTPs: each 200 uM  
Tag Polymerase: 0.3 units  
Total Vol: 10 uL

Buffer:

MgCl2: 1.5mM  
KCl: 50mM  
Tris: 10mM  
pH: 8.3.

Location/Qualifiers

1..524

/organism="Homo sapiens"

FEATURES

source

STS

primer\_bind

primer\_bind 73 a 101 c 110 g 122 t 118 others  
BASE COUNT complement(211..230)

ORIGIN















```

/translation="MEMNRLSTNKHGDTLYLFGAGNAGYALSLIRAEICGP
GALLGDDOIRNVITAAHAFVIFPMWPIHIGFGKMWLPIMLAGDMMAPRNNSF
WLPSPFLLLLASSWAGAGTGVVPPPLAGMLAGASVDLPIISLHAGVSSILG
AINEITITIMKPPAMSOXOTPLEWVSILTAALLSLPLVLAAGITMLDRNLMTT
FEPDAGGADILYQHLFEMFEGHEVYILIPGGMISHTYVYSGKKEPGYGNWMA
MMSIGFGEIVMAHMEFTMGVDYTRAYFSAMITAIPIGVVFSMLATLHGKNTKM
SPAMIANAGIFLFTVGLGLIANSLSLDIYVAYVAHHYHYSMAVAINCG
FYHMPPLFSGYIMDDTMAKIHFTIMFGVMTFFPQHFGLSGMPRISYDPAYITTW
NTVSSMGSTSLTAVMLEWMEAFASKREVAVELTTTNIEMLHGCPPTHYTFEEP
TYVVLK"
trna complement(7817.7887)
/gene="trna-ser"
/notes="anticodon ucn"
complement(7817.12618)
/gene="trna-ser"
7892.7958
/gene="trna-asp"
7892.7958
/gene="trna-asp"
7959.8642
/gene="trna-ser"
/codon_start=1
/transl_table=2
/product="cytochrome c oxidase subunit II"
/protein_id="CA445260.1"
/db_xref="PIR:g13435"
/db_xref="GI:13435"
/db_xref="SWISS-PROT:Q00528"
/translation="MAVPLQMGLODAPSPMEELHHPDHTLIVFLISSLVYITSL
MTTKLHSTNDAGQEVYWTLPALITLILALPSRLIYMDDELINISLTYKTHG
QMTWSTEDYEDLNDSDYTPQELKPGRLLEVDNRVLMDEINIRLISSEVD
HSAVPSLGLKTAIPGRNLQTLMTMRPLTYGQCEISGNSHSEMPYLVLEPLSH
FEWWSYSL"
trna 8644.8712
/gene="trna-lys"
8644.8712
/gene="trna-lys"
8714.8917
/gene="trna-ser"
/codon_start=1
/transl_table=2
/product="ATPase subunit 8"
/protein_id="CA445261.1"
/db_xref="PIR:g13436"
/db_xref="GI:13436"
/db_xref="SWISS-PROT:Q00522"
/translation="MPOLDNSTWLIMLSMILFLITFOLKSKHYFPPTPEKHTPL
LNSAPMEKWKTIYSPSLPIQ"
8875.9555
/gene="trna-ser"
/codon_start=1
/transl_table=2
/product="ATPase subunit 6"
/protein_id="CA445262.1"
/db_xref="PIR:g13437"
/db_xref="GI:13437"
/db_xref="SWISS-PROT:Q00521"
/translation="MNELEASAPATPMGLPIVILVLEPSLIPSPDLINRLAS
IOOWLIQTSKOMLSHNKRGQTWMLMLSLIFISSTNLGLSPSTPTTOLSNL
GMAIPLMAGVITGFRTKTIASLAFELPOSTPLPLMVLITITISLFIQPMALAVL
TANITAGHLIHLIGATTLALMDISATATITFTLILITLEFAVALIOAVYFTLLV
SLYLDHNT"
trna 9555.>10337
/gene="trna-ser"
/notes="no stop codon"

```

Note: remainder of annotations omitted.

Query Match 2.9%; Score 19; DB 23; Length 16826;  
 Best Local Similarity 100.0%; Pred. No. 7.20e+00;  
 Matches 19; Mismatches 0; Indels 0; Gaps 0;  
 Db...3254 CTAATGATCCAAAATT 3272

CP 459 CTAATGATCCAAAATT 441

RESULT 12 CEK08H2 36539 bp DNA INV 23-NOV-1998

LOCUS Caenorhabditis elegans cosmid K08H2, complete sequence.

DEFINITION 270210

ACCSSION 91229075

VERSION 270210.1 GI:1229075

KEYWORDS HTG: Murine modifier 2 protein like; Protein kinase; RNA polymerase

SOURCE tRNA-Tth.

ORGANISM Caenorhabditis elegans.

REFERENCE 1 Rhabditina: Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 2 (bases 1 to 36539)

REFERENCE 3 Submitted (16-MAR-1996) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu

REFERENCE 4 (bases 1 to 36539)

REFERENCE 5 Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,

Boatfield, J., Burton, D., Connell, M., Copsey, T., Cooper, J.,

Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,

Fullon, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,

Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,

Latreille, P., Lightning, J., Lloyd, C., McMurtry, A., Mortimore, B.,

O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A.,

Saunders, D., Snowken, R., Smaldon, N., Smith, A., Sonnenhammer, E.,

Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,

Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,

Wilkinson-Sproat, J., and Wohldman, P.

2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

elegans

Nature 368 (6466), 32-38 (1994)

JOURNAL 94150718

MEDLINE

COMMENT

Coding sequences below are predicted from computer analysis, using

predictions from GeneFinder (P. Green, U. Washington), and other

available information.

For a graphical representation of this sequence and its analysis

see:

http://webace.sanger.ac.uk/cgi-

bin/display?db=wormace&class=Sequence&object=K08H2

Current sequence finishing criteria for the C. elegans genome

sequencing consortium are that all bases are either sequenced

unambiguously on both strands, or on a single strand with both a

dye primer and dye terminator reaction, from distinct subclones.

Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of

the specified clone. It may be shorter because we arrange for a small

overlapping sections once, or longer because we arrange for a small

overlap between neighbouring submissions.

It may be shorter because we only sequence overlapping sections

once, or longer because we arrange for a small overlap between

neighbouring submissions.

The true left end of clone K08H2 is at 1 in this sequence. The true

right end of clone K08H2 is at 7075 in

sequence 250742.

The true left end of clone K09A11 is at 36436 in this sequence. The

true right end of clone R12H7 is at 15206 in this sequence. The

start of this sequence (1..101) overlaps with the end of sequence

250755.

The end of this sequence (36436..36539) overlaps with the start of

sequence 250742.

Location/Qualifiers

1..36539

/organism="Caenorhabditis elegans"

/db\_xref="taxon:6239"

/chromosome="X"

/clone="K08H2"



```
gene complement(1139..1669)
CDS /gene="K08H2.1"
    /note="predicted using GeneFinder; Similarity to Human RNA
    polymerase II elongation factor like protein (Pir Acc. No.
    S55509); cDNA EST EMBL:D72393 comes from this gene; cDNA
    EST EMBL:D75263 comes from this gene; cDNA EST yk374f9.3
    comes from this gene; cDNA EST yk374f9.5 comes from this
    gene"
    /codon_start=1
    /protein_id="CAA94149.1"
    /db_xref="PID:e1347978"
    /db_xref="PID:g3878386"
    /db_xref="GI:3878386"
    /translation="MTVGIQKPVFKIQSCDGVFAVDEWIEQSKCFSVLAKOMNLV
    OPHPLKTSICFVLVLTIEWCYHNRNEDODHQRILTMVDQFIRTSNAILILAEAA
    RLLEIVGLDVACRAVSTMLGRTLNEVLMKLMVGVESDESDSLDEVLELVYDE
    NEDVEIETPISAA"
    complement(9283..9962)
    /gene="K08H2.2"
    /note="complement(join(9283..9576,9654..9962))"
    /codon_start=1
    /protein_id="CAA94148.1"
    /db_xref="PID:e1347977"
    /db_xref="PID:g3878385"
    /db_xref="GI:3878385"
    /translation="MSFSVFEYDESSSQQLIOEAVAASKIMDLVNSSGNGKAKVE
    LAEKMLAAVAASISLYNAIKREEEKLKEEEMRQKRAVEENKLSPELKKAVDL
    NVASIKPHVDVDPDAQPEDIALPTGGLAKCSCSGCTELPOPAYQIINKNEODKRW
    NVRKETRIIVKPPSSAEKVNVTGKKKKSKGKVAKE"
    12427..13369
    /gene="K08H2.3"
    /note="complement(join(12427..12726,12816..13121,13172..13369))"
    /codon_start=1
    /protein_id="CAA94150.1"
    /db_xref="PID:e1347979"
    /db_xref="PID:g3878387"
    /db_xref="GI:3878387"
    /translation="MNGSEKRLHELPTEGRLERRVTEADGIAAFKRMVYMLDKR
    HGELPTRLLMRGILIDGYSINLHMDIKIRVLEKGGVPMEDDKLVYKIRKRF
    ECVTPIFKIARFADMRNDGIPMEYEEDDPICGLEYSERAESEFIELYNTWFHED
    VDQOPGRDEDPAAEQENSEDEVEVERPSVDRLALFLGEMVLTMOPTHVDYPPAPK
    PGDIVIPIGLKLKCCSPGCTRELPOPAVRIIDEAERRRFPDNNKLRK"
    13656..14953
    /gene="K08H2.4"
    /note="complement(join(13656..13700,13828..13911,14596..14685,14744..14953))"
    /codon_start=1
    /protein_id="CAA94151.1"
    /db_xref="PID:e1347980"
    /db_xref="PID:g3878388"
    /db_xref="GI:3878388"
    /translation="MOTALVGOHOCCHDIPOKARMSKQPVSSLDLEITTKVEKC
    PNNSDPEQDEPVOVMSAKIRALVEKMSIQEHTVTPSPKGETIVIPITGSLAK
    CVCSPNCKKELPQPAVLIITEDNNOMHMLKNAKKVYH"
    complement(16114..17283)
    /gene="K08H2.5"
    /note="complement(join(16114..16302,16348..16605,16651..16838,
    16887..17163,17257..17283))"
    /gene="K08H2.5"
    /note="Similarity to Mouse CLK protein kinase

(SW:KICK_MOUSE)"
    /codon_start=1
    /protein_id="CAA94152.1"
    /db_xref="PID:e1347981"
    /db_xref="PID:g3878389"
    /db_xref="GI:3878389"
    /translation="MSFVARSYACPYVDVLRNYKIVELVSGSEYGTAFQVADSR
    HSTLAKNSMNETIFECNCEKLEKERTVQRLSTLTNERSRPTLVDFVDSGLCLV
    MTKGDSISLGVCKRNDKPEKSPVTLKIMLSVGSLOTIHSLSGIHNDIHNNVLFK
    TTPASPCKLIDYGVGKKFRNRGRNVYRNPQRDVNFEKGVSMNVMGVPVLMKD
    FSSMLFGIKTISGTSILETGSVSRHOKMFEELDPSRFLRSVWMLKACVIVDSDA
    EKNYAGVFNAIKOGFPPEDEEILNHSLOSLAVY"
    20433..21840
    /gene="K08H2.6"
    /note="complement(join(20433..20525,20601..20684,20738..20827,20880..21082,
    21756..21840))"
    /codon_start=1
    /protein_id="CAA94153.1"
    /db_xref="PID:e1347982"
    /db_xref="PID:g3878390"
    /db_xref="GI:3878390"
    /translation="MAVLSNVDLGSYIOGIRTOEHSMLGLNRENCSSKWNMSRL
    ENKRTTAGVGNQSEPCDIVOGAFVGMVLRITPENTPTGILCCSQSVSLPVYKSI
    DTARLIFEAELRFTTSOPTROQRILRKVCFSKIMSC"
    26779..31004
    /gene="K08H2.8"
    /note="Similarity to steroid hormone receptors"
    /codon_start=1
    /protein_id="CAA94154.1"
    /db_xref="PID:e1347983"
    /db_xref="PID:g3878391"
    /db_xref="GI:3878391"
    /translation="MNGSEKRLHELPTEGRLERRVTEADGIAAFKRMVYMLDKR
    HGELPTRLLMRGILIDGYSINLHMDIKIRVLEKGGVPMEDDKLVYKIRKRF
    ECVTPIFKIARFADMRNDGIPMEYEEDDPICGLEYSERAESEFIELYNTWFHED
    VDQOPGRDEDPAAEQENSEDEVEVERPSVDRLALFLGEMVLTMOPTHVDYPPAPK
    PGDIVIPIGLKLKCCSPGCTRELPOPAVRIIDEAERRRFPDNNKLRK"
    13656..14953
    /gene="K08H2.4"
    /note="complement(join(13656..13700,13828..13911,14596..14685,14744..14953))"
    /codon_start=1
    /protein_id="CAA94151.1"
    /db_xref="PID:e1347980"
    /db_xref="PID:g3878388"
    /db_xref="GI:3878388"
    /translation="MOTALVGOHOCCHDIPOKARMSKQPVSSLDLEITTKVEKC
    PNNSDPEQDEPVOVMSAKIRALVEKMSIQEHTVTPSPKGETIVIPITGSLAK
    CVCSPNCKKELPQPAVLIITEDNNOMHMLKNAKKVYH"
    complement(16114..17283)
    /gene="K08H2.5"
    /note="complement(join(16114..16302,16348..16605,16651..16838,
    16887..17163,17257..17283))"
    /gene="K08H2.5"
    /note="Similarity to Mouse CLK protein kinase
```







```

gene
  complement(15179..18115)
  /gene="F41D3.5"
  /complement(join(15179..15442,15488..15698,15745..15859,
15910..16036,16093..16412,16877..17043,17092..17293,
17486..17624,17673..17937,17991..18115))
  /gene="F41D3.5"
  /codon_start=1
  /protein_id="CAB04376.1"
  /db_xref="PID:e1346783"
  /db_xref="PID:g3876989"
  /db_xref="GI:3876989"
  /db_xref="SPTREMBL:O45509"
  /translation="MSEASKPKRLDQGLRGALAVLGFHFYPAQFNGYLGVDQF
VLGSCMLKRAEDSTCTLTLLFYSKRFILPLYLILLSMISTYTFPSI
ENOKSATRAMLFVSNAPTSAODNYSMLTKADIDITHTWSLVEQFCLPFIL
ATRSKLOLAYVAILGMYTPOGLINRESEVGDPEKLYIDKENIDAKRSHSVN
YVSYLPLESITITINALPTLSTDFVPFVTLTGFLMISENNMLSNKGLVIGNIS
YSLYLHWPVYVWKLCEGDQDLILALSSVLAVIHEFEKRWYLCSTISGLV
FYLLILNLNFKHEFKMGPKKEVEKNVLANLYDAILKHMOLDQOSICVSC
NEAYAPLGMCHNTSLSPGKYKISVIGSMANHGAMVHEEGSKANTILQGSVASC
EVLVPELVGMCODVDDFOFRMKAENPDYLFITRTYVSGEPPPAITSPDLPIY
QTMKNLEGLLSTKIKITILDAIPRHANNNTKFIEMWKEGVNLYEIDKRLGMEQ
LTDGASYEMARRKTRALVDCDGRVLDYKPFVNSTIKSTRIFDNGSYLTTPKH
LTPRGLEHVRHVKDCATL"
  /complement(18969..21534)
  /gene="F41D3.6"
  /complement(join(18969..19208,19259..19543,19800..19853,
19906..19997,20446..20613,20661..20874,20924..21037,
21084..21244,21298..21405,21468..21534))
  /gene="F41D3.6"
  /note="similar to Vertebrate galactoside-binding lectins"
  /codon_start=1
  /protein_id="CAB04377.1"
  /db_xref="PID:e1346784"
  /db_xref="PID:g3876990"
  /db_xref="GI:3876990"
  /db_xref="SPTREMBL:O45510"
  /translation="MRYFMRYRSSEKSYKLHTLLGLISFVYVNLCLCYDCKQY
LFEPTIYVKGGLQFELVSLGIALRKIRIAVNSDVLQNSLEFLOKIPRI
SPOVLSVPESEPTREFAISDCCREYLSDNHLADSEKSLVIEGHFOSKYFAADML
SKEMNKPEDEPKFRMISTKTSORAKTCVHYRRGDFLIDEGHAGDSYTTISALDHL
KSLHGVITMSNDPRVAVHADHLIDYKDIRIMKTLMEDALIDLHFSQIYCDLVLI
TAPSSIFGWWIIGYLSKNOSAVYVYRDIREFKDQVLOMTKEDYPPITMNLTRGTER
RQVYVNGRPHGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG
PVAOSQTVLVAYVNONILFHPNRPREENVYRSTSGVQWQADRDGSGPFAVGVFN
LTIIVNAGLQWVNNFRGAYRHRNFRAVVSASGANDIGLVVY"
  /complement(22341..22758)
  /gene="F41D3.7"
  /join(22341..22360,22407..22494,22546..22758)
  /gene="F41D3.7"
  /codon_start=1

```

Note: remainder of annotations omitted.

Query Match 2.9%: Score 19; DB 21; Length 39030;

Best Local Similarity 100.0%: Pred. No. 7.20e+00;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1538 TGCTCAATTTTCAAT 1556

Oy 25 TGCTCAATTTTCAAT 43

```

RESULT 14 CELF47D12 41275 bp DNA INV 27-MAR-1995
LOCUS
DEFINITION Caenorhabditis elegans cosmid F47D12.
ACCESSION U22831
NID 9722352
VERSION U22831.1 GI:722352
KEYWORDS
SOURCE Caenorhabditis elegans strain-Bristol N2.

```

```

ORGANISM
  Caenorhabditis elegans
  Eukaryota; Mitochondrial eukaryotes; Metazoa; Nematoda;
  Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
  Rhabditidae; Caenorhabditis.
REFERENCE
  1 (bases 1 to 41275)
  Wilson, R., Ahnscough, R., Anderson, K., Baynes, C., Berks, M.,
  Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
  Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
  Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
  Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
  Latteille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
  O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,
  Saunders, D., Showkhen, R., Smalton, N., Sult, A., Sonnenmer, E.,
  Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
  Vaughan, K., Waterston, R., Watson, A., Wellstock, L.,
  Wilkinson-Sproat, J., and Wohlman, P.
  The C. elegans genome project: contiguous nucleotide sequence of
  over two megabases from chromosome III
  Nature (1994) In press
  2 (bases 1 to 41275)
  Tach, A.
  The sequence of C. elegans cosmid F47D12
  Unpublished (1994)
  3 (bases 1 to 41275)
  Waterston, R.
  Direct Submission
  Submitted (14-MAR-1995)
  Submitted by:
  Genome Sequencing Center
  Department of Genetics, Washington University,
  St. Louis, MO 63110, USA, and
  Sanger Centre, Hinxton Hall
  Cambridge CB10 1RQ, England
  e-mail: r.wellstock@wustl.edu and jesse@sanger.ac.uk

```

#### NEIGHBORING COSMID INFORMATION:

The 5' cosmid is C18P10, 399 bp overlap; 3' cosmid is C31H11, 200 bp overlap. Actual start of this cosmid is at base position 200 of CELF47D12; actual end is at 41076 of CELF47D12

#### NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).

```

FEATURES
  source
    1..41275
    /organism="Caenorhabditis elegans"
    /db_xref="taxon:6239"
    /map="III"
    /strain="Bristol N2"
  gene
    15..949
    /gene="F47D12.4"
    /join(15..251,345..424,607..781,833..949)
    /gene="F47D12.4"
    /note="coded for by C. elegans cDNA CEES77R"
    /codon_start=1
    /protein_id="AAA64316.1"
    /db_xref="PID:g722356"
    /db_xref="GI:722356"
    /translation="MYQATPRDMGKPPVRYGKTSPPGFVRCYCEHKKYNNVOYV
ELSKCSKKTKTWDDKRRPYELAQAQAEYQAEVAYLGGEDAMRRKRRKKPPHAP
KRAUSAFFYISQDKRPEIOAGHPDKVQVQAEQKMKVLPQETKDYEQKQADKD
RYADEMRNYKAEQMSGDHYYDDNHHVHVVDINSQNS"
    /complement(1459..4677)
    /gene="F47D12.5"
    /complement(join(1459..1834,2647..3255,3416..3704,
3766..3923,3975..4240,4367..4488,4596..4677))
    /gene="F47D12.5"

```



/note="similar to C. elegans early embryogenesis zyg-11 protein; coded for by C. elegans cDNA CEMS12F"

by C. elegans cDNA CEMS12F"

/codon\_start=1

/protein\_id="AA64317.1"

/db\_xref="pid:g722357"

/db\_xref="gi:722357"

/translation="MWLKRHCLESLIGDLDYIGYRDLIDNTIDIFELLKALLN  
ETRNHLHYDIDKSHAFSKKMTNIGTILPVLSLITCGRHLOSDDFOOLCESIPMLT  
YDISETNPINISLKLKLESLIMRSELECFQFDLNLGLRLIDVSRSMKH  
DKILNVOYOSNKVKELRPLESGDINKNLEIOLNSHDNLQOYVALNCHIRSL  
HHEVFLNATLTLSSLSKSLHYGVRISSEVERLENMLHPKIPDPDNDRESEC  
VATINSALKIESLIGIGLCLMHLARNDARQAPDIOVLCKLDGVNNSRL  
FLATKDFNISLWETISLKLKLESSLKPMSEVERCRAAGCISRNVEDDEKTRALPI  
FAYIKRISOSQYVYIDFNHKLMPKIIHLKKSIELEKTYWIFPCDLITLQNSAD  
ACEKLACEMANVLECFGNLTKASDHKLKILKNVLELTPKYNKHYADARISVL  
OERLWDEDEARVYFTVALIISNFTYINREPAISRECEPESINVIKIPVSOK  
IYKNGIKILKLSRSDSLMALVITKSMQDSFVNLTTKSNLNTNTEIIMN  
MKAELKNLT"

complement(5714..6261)

/gene="F47D12.6"

/join(5714..5842,5885..6011,6215..6261)

/gene="F47D12.6"

/codon\_start=1

/evidence=not\_experimental

/protein\_id="AA64318.1"

/db\_xref="pid:g722358"

/db\_xref="gi:722358"

/translation="MDNIOFEPDOYVHLPESSKHDHAYIILSVIILILLICN  
LCICYRIRORRELIDYPSNTQYIPFRNVRKPYRTESGTSNNRMLPPQHV"

complement(12410..14864)

/gene="F47D12.7"

/join(12410..12529,12577..12871,13010..13210,  
13277..13499,13545..13635,13681..13834,13881..13979,  
14028..14116,14167..14251,14291..14393,14440..14534,  
14679..14864)

/gene="F47D12.7"

/note="similar to C. elegans protein T16H12.6 throughout  
entire protein; weakly similar to M. musculus MIP protein  
(sp:MIP\_MOUSE.p28575)"

/codon\_start=1

/evidence=not\_experimental

/protein\_id="AA64319.1"

/db\_xref="pid:g722359"

/db\_xref="gi:722359"

/translation="MDICDISASOLATSPPEPTAIDPNSIIEESPYKRVKFSPELKS  
FEETLSPTVIVRNNEEYFDRLSLFYSNRYVLFSSKFRDSTTHRIIRLSAD  
LHLLIIPKAFEGGKIPNTILOKALIELLEPAFLQMSIALDIYSDICKNLHENTIK  
IRLALLHTTLAVRWSVRKFTLFATNVLSLKENELIGLTDKHLNKSDET  
TVVNVNIRKNSPLQSDRLQFAQRNFSRPOPATRYEYIRKIPMDALVCEGMSR  
GVAOKIEVNTSRSDMOTCNFYDIPNIRAYGIVEDEKLIYVGFDIKQFOTV  
LEPLSTKEWRGRGANNMDCRYTSARVNISYGRVLAECGGMVSVRLTAEMYDRA  
DOMTEVANNOMSDGAVYTDINKIYVIGGFGDRNHOGEVYDPYDLNHPSSMR  
TRRTGCTAVSINKOVCMITIGGFNGNRRLDSAEYIDRBEGLMHPPTLQTRKSFSSC  
MDICVYVAGGFDGQTTKESERLDRSKWQALPMAEKSKALRWVTLSDHPELDEL  
FDIPDTGIVTSW"

17507..18339

/gene="F47D12.3"

/join(17507..17746,18157..18339)

/gene="F47D12.3"

/codon\_start=1

/evidence=not\_experimental

/protein\_id="AA64315.1"

/db\_xref="pid:g722355"

/db\_xref="gi:722355"

/translation="WYNELSFNCEDAMNLFVGGEMYPVQVKTLMNPTGSGYRQVY  
KVSDAIKVGVQWMDAPNHIKFRVIDIDRGVLFVRLVQLYLRNGKLTSLPDDITLES  
LVAAEFEGLEKREMLKIKMLTKGRQYACCEBDS"

complement(18812..21512)

/gene="F47D12.9"

/join(18812..19005,19378..19946,20005..20244,  
20307..20365,21091..21375,21422..21512)

/gene="F47D12.9"

/note="coded for by C. elegans cDNA CEMS28R; coded for by  
C. elegans cDNA CEMS16R; coded for by C. elegans cDNA  
CEMS16F"

/codon\_start=1

/protein\_id="AA64320.1"

/db\_xref="pid:g722360"

/db\_xref="gi:722360"

/translation="MSRFRKARSNMFGONDSRWPEPPVRLNLTATPEEPLP  
KLEDOPEEGPLNMGFMHPRTKYKMTOPMPQFSKSDLDMEKARAKFOAN  
RRFTSGSFIOPRVEPFTITLMDLILGCTMARVRRHIESLNCNKPSTIKTP  
IRHYVSGCEPFIQVSETPGRIYGTIVNPNVGAARSAVVEVDISIGDTIOSSSR  
EAYOLPIRSRNSNAGFNTLGLTPRMLEDDEGSDPSLIDYAVTRNSFTYDQTLAR  
VADVTCMLTVIANDTITRNGVNCSTYVHLEPLAELSDPEAMPTLNSPIYKSWRK  
GNISVGWNAQMSITFEGSCFRENLTDRSFLMSRKNVNLNCSADNLVYMG  
LNDNITKSDLLMNRDHLITGOLNACNTFFVLEKTRPCEVTEGFSIIRIWDFRW  
PKNPMHEHNSNCRNLRLVFPDEKREFFVFAAGSDGVYRGSVLSIGDMICSVKTPHS  
NPIFPRAVYSDCWGRGPNASIIIMVGDSEMRVHSLEL"

28435..34437

/gene="F47D12.2"

/join(28435..28590,33210..33364,33413..33479,33719..33940,  
34318..34437)

/gene="F47D12.2"

/note="similar to muscarinic acetylcholine receptor,  
serotonin receptor, and other G-protein coupled receptors"

/codon\_start=1

/evidence=not\_experimental

/protein\_id="AA64314.1"

/db\_xref="pid:g722354"

/db\_xref="gi:722354"

/translation="MAVASVLALEMLFISTVYVGNLAVLSIYLDKNIROPTNFI  
FLAISDLILGLEGIPVYTAFLNNEMWLGVDLQMLSDIYICVLASITYVLGITV  
DRYSVKRPATYRNMTPOGVLLIIFIMLVPSIIFSVISFGYGTGTGRILKETECC  
VYOFMTNPLNMGMYISYWTLEFVLLYVWGIYRAKRLAKSDQKTRRLALTEMR  
RBEVSRTSDACNSSSDSPNDNSNSK"

36832..39617

/gene="F47D12.1"

/join(36832..36896,37155..37319,37401..37692,38052..38123,  
38345..38629,39333..39617)

/gene="F47D12.1"

/codon\_start=1

/evidence=not\_experimental

/protein\_id="AA64313.1"

/db\_xref="pid:g722353"

/db\_xref="gi:722353"

/translation="MDFTKDNELVPRPTPDNDTYSNPNFMSISEQLTNGFSROEPS  
VIERESTACVPEPSHASLENNENHAAHKPELSTPFIADSVSSVWGDDILARA  
MSRISRSTSMGTARATYVIEIVENLEALIKCNLELREDEKNEBKXNGLEN

\*\*\*  
Note: remainder of annotations omitted.

Query Match 2.9%; Score 19; DB 21; Length 41275;  
Best Local Similarity 100.0%; Pred. No. 7.20e+00;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 6287 AACGATACAGATTAAAT 6305  
|||||  
Oy 478 AACGATACAGATTAAAT 496

RESULT 15  
LOCUS AB018112 70475 bp DNA PLN 20-FEB-1999  
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MAB16,  
complete sequence.  
ACCESSION AB018112  
VERSION 93702730  
KEYWORDS AB018112.1 GI:3702730  
SOURCE HTG.  
Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui P1  
clone:MAB16  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Rosidae; Capnareales; Brassicaceae; Arabidopsis.



REFERENCE 1 (bases 1 to 70475)  
 AUTHORS Nakamura, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-OCT-1998) to the DDBJ/EMBL/Genbank databases.  
 Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of  
 Gene Structure 2, 1532-3, Yana, Kisarazu, Chiba 292, Japan  
 (E-mail: ynakamura@kazusa.or.jp, Tel: +81-438-52-3935,  
 Fax: +81-438-52-3934)

REFERENCE 2 (sites)  
 AUTHORS Nakamura, Y.  
 TITLE Structural Analysis of Arabidopsis thaliana Chromosome 5. IX  
 JOURNAL Unpublished (1998)  
 FEATURES  
 source  
 1. /0475  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="5"  
 /clone="MAB16"  
 /clone\_lib="Mitsui P1"  
 /clone\_1b="11571 g 23373 t  
 11572 c 11571 g 23373 t

BASE COUNT 23859 a 11672 c 11571 g 23373 t  
 ORIGIN

Query Match 2.9% Score 19; DB 27; Length 70475;  
 Best Local Similarity 100.0%; Pred. No. 7.20e+00;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 56397 AATTAAATGAAGATTCCTATT 56415  
 |||  
 QY 520 AATTAAATGAAGATTCCTATT 538

Search completed: Sat Nov 27 14:37:49 1999  
 Job time : 1229 secs.







AUTHORS Zhu,T., Mo,H., Wang,N., Nam,D.S., Cao,Y., Koup,R.A. and Ho,D.D.  
TITLE Genotypic and phenotypic characterization of HIV-1 in patients  
JOURNAL Science 261, 1179-1181 (1993)  
STANDARD full staff review  
COMMENT This sample has been taken from patient F, one of the seven patients (A,M,L,F,C,V,R), in the study(1). All patients had acute, self-limited symptomatic illness with measurable viremia followed by seroconversion. Patient F was infected by patient M via sexual transmission. The authors report that the sequences in this study are found to be uniformly macrophage-tropic and nonsyncytium-inducing. The sequence for sample F clone 1, along with other patient F clones 2-15, and 17, is shown in alignment following the printed text entry. These env sequences cluster with HIV-1 B subtype sequences. See also L21224-L21591 and L24161, L24162.

## FEATURES

CDS

<1..>639  
Location/Qualifiers  
/product="envelope glycoprotein"  
/gene="env"  
/codon\_start=1  
/translation="VSTHLLNGSLAEVEVIRSKNFTNTKIIIVQLNETVEINCTR  
PSNNTSRSHMGCGAFYTTGDIIDIRQACHISREKNNNTLKIVTKLRQYENKI  
IIFKPSGGDPEIVHSGNGGFEYCNTOLENSAMNGEELKSKSGDITLPCRI  
KQFINMGEVGRAMVAPPIRGQIRCSNTITGLITRDGNNNGNNTFRPGGG"

source

<1..>639  
/organism="Human immunodeficiency virus type 1"  
/sequenced\_mol="RNA"

BASE COUNT 266 a 266 c 125 g 150 t  
ORIGIN 720 bp downstream from beginning of env cds

Query Match 2.4%; Score 16; DB 3; Length 639;  
Best Local Similarity 100.0%; Pred. No. 7.65e-02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 56 gatccaaaatttcac 71  
|||||  
CP 453 gatccaaaatttcac 438

RESULT 3  
LOCUS HIV6S2 639 bp ss-RNA VRL 09-SEP-1993  
DEFINITION Human immunodeficiency virus type 1, clone 6, ADC sample M, partial  
SEGMENT env cds, c2v5 region.  
L21505  
2 of 4  
SOURCE Human immunodeficiency virus type 1 (HIV-1), clone 6, ADC sample M.  
ORGANISM Human immunodeficiency virus type 1  
Virus: ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
1 (bases 1 to 639)

REFERENCE  
AUTHORS Zhu,T., Mo,H., Wang,N., Nam,D.S., Cao,Y., Koup,R.A. and Ho,D.D.  
TITLE Genotypic and phenotypic characterization of HIV-1 in patients  
JOURNAL Science 261, 1179-1181 (1993)  
STANDARD full staff review  
COMMENT This sample has been taken from patient M, one of the seven patients (A,M,L,F,C,V,R), in the study(1). All patients had acute, self-limited symptomatic illness with measurable viremia followed by seroconversion. Patient M was a chronically infected male who subsequently transmitted the virus to patient F. The authors report that the sequences in this study are found to be uniformly macrophage-tropic and non-syncytium-inducing. The sequence for sample M clone 6, along with other patient M clones 1, 2, 7-9, 11, 12, 14, 15, 17-23, and 32-34, is shown in alignment following the printed text entry. These env sequences cluster with HIV-1 B subtype sequences. See also L21224-L21591 and L24161, L24162.

## FEATURES

CDS

<1..>639  
Location/Qualifiers  
/product="envelope glycoprotein"  
/gene="env"  
/codon\_start=1  
/translation="VSTHLLNGSLAEVEVIRSKNFTNTKIIIVQLNETVEINCTR  
PSNNTSRSHMGCGAFYTTGDIIDIRQACHISREKNNNTLKIVTKLRQYENKI

source

BASE COUNT 266 a 103 c 123 g 147 t  
ORIGIN 778 bp downstream from beginning of env cds

Query Match 2.4%; Score 16; DB 3; Length 639;  
Best Local Similarity 100.0%; Pred. No. 7.65e-02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 56 gatccaaaatttcac 71  
|||||  
CP 453 gatccaaaatttcac 438

## RESULT 4

LOCUS SIVSABIC 10036 bp ss-RNA VRL 25-MAY-1994  
DEFINITION Simian immunodeficiency virus (SIVagm) complete genomic sequence,  
African green monkey isolate SIVagmSAB-1.  
ACCESSION U04005  
SOURCE Simian immunodeficiency virus PCR-amplified DNA extracted from  
cocultured PBMCs derived from a wild female asymptomatic sabaens  
African green monkey captured in its natural habitat of Senegal.  
ORGANISM Simian immunodeficiency virus  
Virus: ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
1 (bases 1 to 10036)

REFERENCE  
AUTHORS Jin,M.U., Hui,H., Robertson,D.L., Mueller,M.C., Barre-Sinoussi,F.,  
Hirsch,V.M., Allen,J.S., Shaw,G.M., Sharp,P.M. and Hahn,B.H.  
Mosaic Genome Structure of Simian Immunodeficiency Virus from West  
African Green Monkeys  
EMBO J. 13, 2935-2947 (1994)

JOURNAL full staff review  
STANDARD This sequence was obtained as part of a study designed to determine  
the breadth of genetic variation of SIV infecting African green  
monkeys. With this aim in mind, Jin et al. PCR-amplified, cloned  
and sequenced the LTR and partial envelope region from 12 isolates  
(ver-1, ver-2, gri-2, gri-3, sab-1, sab-2, sab-3, sab-4, tan-1,  
tan-17, tan-40, and tan-49). These isolates were derived from  
different African green monkey species inhabiting various  
geographical regions.  
Virus sab-1 was derived from a wild female asymptomatic sabaens  
African green monkey. She was caught in her natural habitat of  
Senegal and subsequently exported to the United States. From the  
point of capture and throughout the study, she was housed in an  
individual cage.  
Phylogenetic analysis suggests that each AGM species is infected  
with a distinct subtype of SIVagm. Interestingly, the sabaens viral  
genome seems to be a mosaic of many viruses infecting different  
primate species [1]. Virus sab-1 forms a distinct lineage in the 5'  
end of gag, clusters closest to the HIV2/SIVsm group in the 3' gag  
and 5' pol regions, and clusters closest to the other SIVagm  
lineages in the 3' pol, env and nef regions [1]. This feature  
was demonstrated for two viral sabaens strains, sab-1 and sab-D37  
(sequenced only over the 5' end of the pol gene). Jin et al.  
conclude that an ancestor of the sabaens virus was a recombinant  
with two breakpoints: one roughly at codon 310 of gag and the other  
at 710 of pol.

NCBI g1: 466229

## FEATURES

source

Location/Qualifiers  
1..10036  
/clone="SAB-1/M78"  
/strain="SIVagmSAB-1"  
/organism="Simian immunodeficiency virus"  
/specific\_host="African green monkey"  
/proviral  
/note="sabaens subtype"  
1..767  
481..527  
LTR  
LTR  
/note="putative TAR 1"



LTR  
528..568  
/note="putative TAR 2"  
1017..2681  
/note="NCBI gi: 466230"  
/codon\_start=1  
/product="Gag polyprotein"  
/translation="MGASNSVLSGRKLDAFESVRLPNPKKKYKRLHLMWASKELDRF  
SLSANLEKEGVKILSVLLPLPTGSENLALFELCCVLAICHAIEIVKDEEKA  
KKEEVPAMETASATISGQRELOAKKNEPTVPSGSRNYPPIVSNOMVHOPPL  
SRRLNMYKVIIEEKESAEVPMPSALDEGAPIDINOMLNAGVHOGALIDVMDI  
NEADMDLRHPPOOPPAOVLDRPOGSDIAGTSTIOEJEMTTBRANAYNAGIT  
KMWILGKCYAMINPVLIDKQKPEPEVDYRETKALAEETDRAVNMWMTOS  
LLIQANPDCVTYKGLGNPTELEMLTQOIGGQKARLAEAMTAFOQOYVGN  
IFVOGARGRPLGGRGRPLNPKYNGKPGHARFCAPARQSCMRCSPDHQK  
DOKQVNFIFGFGMGKRPNEPLTISIRTPAMERDYPREEMVYADNPPRGPPD  
DPATLLKQYAVGKRQKQONHSPQSPYEAVSSLSLFGEO"  
<2387..5506  
/note="N-terminal uncertain; NCBI gi: 466231"  
/codon\_start=1  
/product="POL protein"  
/translation="FPRVPLGROETPEPSDLHGTNSPANGTGLQOAGKLVCRQTS  
DQTRARRSSNPVAKVCCSGEAEFAVAKPLATTEPRLGLQLPQVSLMRPMKTV  
TEGQVTLDDGADSVQGLEGNMKPRITIGGICINIKAYHNOGVKTEDTK  
ATLVGETPVNIGRNYLQGLTYLTLQRETEPRLVHLKPODGFRIKQMPSEKRI  
EAKACEDLEKQGLERIGPENPNTVPFALRKDKTQWRLMDFROLKSTQDEQE  
VOLGIPAGLOQREQIVLDIGDAFSCPLDPDFQKTYAFTIPVNNNEPGRIOYK  
VLPQWKSPTIFQITANKILOERQKNDVDIYQMDMLASDPKAEHVMYQOL  
RUYLETGKTEPEKFKQDPPLMNGELYPKKMOLEITPEREEMVYADNIOKLGK  
LMAQIYNGITKTKLRLRGARPLETVCWTEPELELENROLIRKQOQYVDP  
ALPRAVKLKGDMQIOIYOPNKILVGVAKIKTINELRLAGLVQIGKES  
IYIWQIPIMLPEVERLEWQWMSDYMOTVLPMEWVSTPOLIRLWYLVADPGE  
AVYVYGANRNSKEGAGYLTDRDQKVALENTQAEALILALRDSKNI  
INDSQAMGIIAGEPTESDNNITVQIIEELKEAYILAMPARKGVGENEEDKVS  
OGIRQVLEDRIEAQEEDHKHANRMSQOEGEPAIYAKELVACAPCOIKGESV  
GVDAVSPGWMDCTHLEGGKILIVHVASGTEAPVIOENKCAAHILIASMP  
ITQHTDNTNTSOQVAHICWAKIEHTFGVPYNQSGVYESNNKQLEITQND  
DAERLETAIVMAVHINFRKGIIGYSAERLINIHELETKILOQKISIKQNRV  
YREGRDPAWKPAKLIMKGEAVVIOEGEELTIPRRRAKIKIDYKALDSQAPLEG  
NGRTAGEVD"  
5439..6143  
/note="NCBI gi: 466232"  
/codon\_start=1  
/product="VIF protein"  
/translation="MEKHWIVPLMKVTGTGOERMTSLVKYHMHVSKOCVHNRYPHT  
KIRMYVSYQEWVILPKDGLIKVTNYMLTEPEKLEIYAGIGSKEMTEIDPM  
TADHITHSVPCQDRAVOAALRGEKYMCKHOGHOPDPSLOALAYVNGL  
RVVAPSRSGSGSQESQRRDTRAMRMGRQAVRMAFRHVTGPQFRGVPPLPK  
ESPPSLVEYCGTSH"  
5911..6333  
/note="NCBI gi: 466235"  
/codon\_start=1  
/product="VPX protein"  
/translation="MASGWLPPVYGDPKDKPKNREIIPGMLTEWDLPREPDEML  
RMLQDLNSEAOCHPERNLFLRMNVIPEVHIDHQTLEWYKTCRILQALFVHM  
KORCKPKTHPVPAGAPPPGLGASGASAAAGL"  
6254..6475  
/note="NCBI gi: 466233"  
/codon\_start=1  
/product="TAR protein"  
/translation="MDQEQARPQWELEQELHRLQACDNTCFCKVCCFHCICFH  
joIn(6411..6474,8683..8918)  
/note="NCBI gi: 466234"  
/codon\_start=1  
/product="REV protein"  
/translation="MSLQDELLRRRIILKFLYTNPPYGGGTARQRRARQRMWAKO  
ROQVHLAERRIETPVSQIDHLEQEDQVLDNLQOPSLPFGHPTENOTANSSS"  
6481..8679  
/note="NCBI gi: 466236"  
/codon\_start=1  
/product="ENV protein"  
/translation="MKLLTVLLMLSGCWSLWMLVOYVTVFYGIPIVKNSSVOAFCKTP  
NNLMASINCIQDDEPQIAEVPPIPIITEKEDAMKRNRPDLQOAESNIHLLESTLK

PCVKLSMPCIKMNCYRLGGAATTSPTSTARPEVSVGENDSVIEOMEKEQAMNC  
SEFAMGVRDVKKNSTVMDDQEVVCEEGREKSNATHVGCYMIHCNTSVLKACDKT  
YWDTRLRVACAPGALLRCADDTSGHAKCAGNVAUSACTRLINTVTSIGINGSY  
AMRETIKONGNSDVSILRLNRYNLIIRGRPNKQVLPVLTINAGIYFVSOKNTR  
LKQANCWTFGKMRGAKWEKETIYALPKKTSIGTIDTKIFLQRMGCPSEFFPNC  
QGEFFYCKMDVLANLNNKSVDPDNNCAKNNTKPCWORTYVCHIROYVNDWYLSK  
KIVAPPREGHLECNSTATALVYELVNSKRNNTVLSPOIESIMANELGDYKLVKIP  
IGFATPKRYRYGPERQKRVPEVLGFLGAGAMAAATLALVQSQOLAGLLOQ  
OKNLAABOQOQMKLTIWGMKLNAAVTLLEKYLEQOARINIMGCAFROYCHTTL  
WKYNTPEMNEANTQEMEROLEKYEYANISRLIEQAHEDQKILDSYORLYMSDWSM  
FDLTFWCKMKAIVAGIYARVLVILGILKFRKGYADLSLPESH"  
8921..9601  
/note="NCBI gi: 466237"  
/codon\_start=1  
/product="NEF protein"  
/translation="MGKSSKQOQSHSLIMSKLROAPVIOYDMLADPLQSSHIOE  
EKAQLRBLQIQCQSSRTREBYKHKHGRQPSWYDEDEEVGFVPRCLPLRANITYK  
LAIDGHLERKGGLEGLEYSSRRKKILDLALNWMGLVDQMONTDLPGRTPYPCFG  
WCFKLVPVDSLSEAKNSENHCLHRAQYAVEDDAWKETLWMKFDPLADVAVMRLHP  
EVPASAQ"  
9270..10036  
LTR  
BASE COUNT 3293 a. 1975 c 2517 g 2251 t  
ORIGIN  
Query Match 2.4%; Score 16; DB 3; Length 10036;  
Best Local Similarity 100.0%; Pred. No. 7.65e-02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 3712 tggatcacaata 3727  
|||||  
QY 589 TGATATTCAAAAATTA 604  
RESULT 5  
LOCUS HIV169 1459 bp ss-RNA VRL 20-APR-1993  
DEFINITION Human immunodeficiency virus type 1, isolate v169 taken from a  
Rwandan national residing in Belgium, gag region.  
ACCESSION L11796  
SOURCE Human immunodeficiency virus type 1 (HIV-1), Rwandan isolate v169.  
ORGANISM Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
REFERENCE  
1 (bases 1 to 1459)  
Louwaagle,J.J., McCutchan,F., Brennan,T., Peeters,M., Brennan,T.,  
Sanders-Buell,E., Eddy,G., van der Groen,G., Franssen,K.,  
Gershy-Damet,M., Deleys,R. and Burke,D.  
Phylogenetic analysis of gag genes from seventy international HIV-1  
isolates provides evidence for multiple genotypes  
AIDS 7, 769-780 (1993)  
JOURNAL  
STANDARD  
COMMENT  
Kindly provided prior to publication by Henry M. Jackson Foundation  
Research Laboratory, 1500 East Gude Drive, Rockville, MD 20850.  
The authors group the outlying V169 gag sequence with their  
subtype E. The naming of this subsequence was made independently of  
envelope subtype studies, and so gag sequences in this subtype  
are not necessarily related to envelope sequences classified as  
subtype E in the 1997 and 1993 compendiums.  
location/Qualifiers  
1..1459  
/product="gag protein"  
/gene="gag"  
/codon\_start=1  
/translation="MGARASVLSGKRLDAFESVRLPNPKKKYKRLHLMWASKELDRF  
ALDRPLERSBCKRIITIGLOLSDTGSEMLRNLTAVALYFYQKYVDYKREALE  
KLEEQNSQOKKQVADKGVSONYPIVONLQGVHQAISPTLNAEMDVYIEKAF  
SPEVIMFSALEGATPDLNMLNTVGGHQAAMQMLKDTINEAEAMDRLHPVAGP  
NPPGMRERPSGSDIGTSTLOEIOYMTSNPIVGDYIKRWIILGLNKIYRMSPV  
SLIDIROGKEPFDYVDFEKLTAEEATROKGMMDTILLIONANDCKTILKAMG  
PGATLEEMTACQGVGGRKRVLAETMSQATNAIMQSNFNGQRRTYCFNCGK  
EGHVAKNCAPRKKCKMCKGREGHQMKQDTEKQANFLGKIMPSNKGRENFLQSRPEP  
SAPPAESFGFREITIPSPKQEKDGLSPPLASIKSX"  
1..1459  
/organism="Human immunodeficiency virus type 1"



```

/isolate="VI69"
/cell_type="lymphocyte"
/proviral
/sequenced_mol="DNA"
/tissue_type="blood"
BASE COUNT      531 a      277 c      365 g      286 t
ORIGIN      gag cds start
Query Match      2.3%; Score 15; DB 3; Length 1459;
Best Local Similarity 100.0%; Pred. No. 7.01e-01;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      300 tttagagaagctaga 314
        |||||||
        163 tttagagaagctaga 177

RESULT      6      HIYDJ373A      3340 bp ds-DNA      VRL      31-DEC-1993
LOCUS      Human immunodeficiency virus type 1, DJ373 from Djibouti,
DEFINITION      proviral DNA encoding env, tat, vpu, rev, and nef genes.
ACCESSION      L23065
KEYWORDS      env gene; envelope glycoprotein; nef gene; rev gene; tat gene;
SOURCE      vpu gene.
              Human immunodeficiency virus type 1 (HIV-1), DJ273 from Djibouti,
              PCR amplification of proviral DNA from cocultivated PBMC's
              Human immunodeficiency virus type 1
              Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
              Retroviridae; Lentivirinae.
              1 (bases 1 to 3340)
              Louwagie, J., Janssens, W., Mascola, J.J., Fischer, C.L., Hegerich, P.A.,
              van der Groen, G., McCutchan, F.F.E., Eddy, G. and Burke, D.
              Genetic diversity of the HIV-1 envelope glycoprotein
              Unpublished (1993)
              full automatic
COMMENT      These sequences were kindly provided prior to publication by Dr.
              Joost Louwagie of the Henry M. Jackson Foundation Research Laboratory
              Rockville, Maryland. Twenty-one full length gp160 coding sequences
              from eight African countries (Djibouti, Gabon, Kenya, Senegal,
              Somalia, Uganda, Zaire, and Zambia) were sequenced and analyzed with
              thirty-two previously published full-length gp160 env sequences.
              Sequences DJ258, DJ259, K124, SE365, SM45, UG266, UG268, UG274,
              VI191, and VI525 have corresponding gag sequences published by
              Louwagie et. al. in AIDS 7, 769-780 (1993) and in the 1993 compen-
              dium. This sequence clusters with C subtype env sequences. See also
              accession numbers L22939-L22957 and L23064.
              NCBI gi: 437346
FEATURES
SOURCE      Location/Qualifiers
              1..3340
              /organism="Human immunodeficiency virus type 1"
              /isolate="Djibouti"
              /cell_type="lymphocyte"
              /proviral
              /sequenced_mol="DNA"
              /tissue_type="blood"
              join(87..>302,2619..>2707)
              /note="tat protein, exon 2 (first expressed exon)"
              /translation="MEVDPNLEPMNHGSGOPNACTKCYCKKSGHYLVCFOGKGL
              GISYGRKKRRQRRSAPSSSEDHONLSKO"
              /note="tat protein, exon 3 (AA at 2620)"
              /translation="PLSKTQGDPTGPEESKSKKVESKTKADPLD"
              join(226..>302,2619..>2838)
              /note="rev protein, exon 2 (first expressed exon)"
              /translation="WAGRGSDSEALQAVRIKLKYS"
              /note="rev protein, exon 3 (AA at 2621)"
              /translation="SPYEPKGTROAQNRNRMRAROKQIHISISERTLSTCLGRPA
              EPPFQLEPPIRLNDCSESGTSGTEGVN"
              325..>582
              /note="vpu protein"
              /translation="MIDLAKVDYRLAFAFIATVVTIATVIEYRKLNQK
              IDRLIRTERAEADSGNSDGDITELSTVDRGNRLDAVDV
              503..>3031
              /note="env glycoprotein"

```

```

/translation="MRYVGIOMNQOMWIGGILGFPMMLMKNMGNIWTVYGVV
WOEANPLFCASDAKAYETEYHVMWATHACVPTDPSQOELMEVNTENFMKNGM
DOMHODIISLMDDEGLKPCVKTLEPLCVLGNKNVTANGSVTNSTEKDRNCSFN
ITTELDRKOKKRYALFPKLDIVPGLNCSNENFRLNCSMTSTOCAPVSPEDIIHY
CAPAGYALIKCRDTRKNGTGCKRVSVYQCCHGIRKPVYSTOLLNGSTAEEDIIIRS
ENLTNNAKIIIVOLNOSVEINCTRPNNNRQSIIRIGQYFATGDIIGDROAHN
ISROKMETLQOVKGLKEHFNKTIKFPASSGSDIEITTHSGNGGEFFCNTSALF
NDSITNAPINNSSTDANTLLQCRIKQIIMMOEVRAMVAPPIQGIACKSNITGL
ILRDGNSNTEDETRPGGDRDNRSELYKVEYKVEIKPLGATKARRVREKER
AALAAVLGFGAGSGTMSGAISTITQARSLSGTYOOSNLKAEVHOMOLT
VMGKIQJOTRLAERILKQDLIGGCSKILCTTAVPRNSSMSKSOEETIDNM
TWGQWDEINNYETIRLEVSQTOEQDNKDLALDKMQLMSWFDITWMIYIK
IFEMI VGGILGRLIFAVLSIVNRQGYSPLSQTLIPNRPGRDPRGIEEGEG
DKGRSTRIVSGFLALMDLRLSLFESYHRRLDILVARVVELLGGORGETLKYLG
SLVOYWGIELEKSAVSILDTAIVAVGSTRIBIIRIWRNACFNIRRIQGEAA
LO"
CDS      3036..>3340
          /note="nef protein"
          /translation="MGNKWSGWPDAVERIRKTPAAREVAAAEGVGAASODIDKY
          GALTSTSNTPANNADCAWLETOEEEEVGFVPRPQVPLRPMTYKGAFLSFEFLKEKG
          L"
BASE COUNT      1191 a      568 c      791 g      790 t
ORIGIN      226 bp upstream from the beginning of rev cds start
Query Match      2.3%; Score 15; DB 3; Length 3340;
Best Local Similarity 100.0%; Pred. No. 7.01e-01;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      570 tgctttgatgctga 584
        |||||||
        58 tgctttgatgctga 72

RESULT      7      HIYHAN      9076 bp ss-RNA      UNA      20-DEC-1989
LOCUS      Human immunodeficiency virus type 1, isolate HAN-2; complete
DEFINITION      genome.
SOURCE      HIV-1, isolate HAN-2, clones 2/2 and 2/3; infectious proviral DNA.
REFERENCE      1 (bases 1 to 9076)
AUTHORS      Saemann, U., Schneider, J., Mous, J., Brunchorst, U., Schedel, I.,
              Jentsch, K.D. and Hunsman, G.
              Molecular cloning and characterization of a German HIV-1 isolate
              AIDS Res. and Hum. Retroviruses submitted, February 1990
              JOURNAL
              STANDARD
              COMMENT      Kindly submitted in computer readable form by F. Kirchhoff, DPZ,
              Göttingen, 2/2/90.
FEATURES
              Clone HAN2/3, from which most of the sequence shown was derived,
              is inactive perhaps as the result of mutations leading to
              frameshifts in the gp41 portion of the env cds. Clone 2/2 is
              infectious (in MT-2 cells) and its env sequence has been inserted
              in this entry: bases 7100-7608 (annotated below).
              from to/span      description
              259      1758      gag polyprotein
              259      4562      pol polyprotein
              4507      5085      vif protein
              5025      5315      vpr protein
              5286      5510      tat, exon 2 (first expressed exon)
              7841      7931      tat, exon 3 (AA at 7842)
              5435      5510      rev, exon 2 (first expressed exon)
              7841      8115      rev, exon 3 (AA at 7843)
              5527      5772      vpu protein
              5690      8257      env polyprotein
              7283      2      cgc in 2/2; c in 2/3
              7293      3      ggc in 2/2; c in 2/3
              7325      2      gg in 2/2; g in 2/3
              7364      1      t in 2/2; a in 2/3
              7494      1      t in 2/2; c in 2/3
              8259      8861      nef protein
              3276 a      1595 c      2186 g      2019 t
BASE COUNT      3276 a      1595 c      2186 g      2019 t
ORIGIN      5' terminus of LTR U5 region
Query Match      2.3%; Score 15; DB 2; Length 9076;

```



Best Local Similarity 100.0%; Pred. No. 7.01e-01;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 5210 aatacgcacaaact 5224  
|||||  
CP 354 AATACGTCAACAACT 340

RESULT 8 FIVPPR 9468 bp ss-RNA VRL 23-JUL-1990  
LOCUS Feline immunodeficiency virus, PPR isolate, complete genome;  
DEFINITION (San Diego isolate).  
ACCESSION M36968  
SOURCE Feline immunodeficiency virus, clone PPR, genomic DNA. Infectious clone.  
REFERENCE 1 (bases 1 to 9468)  
AUTHORS Phillips,T.R., Talbott,R.L., Lamont,C., Muir,S., Lovelace,K. and Elder,J.H.  
TITLE Comparison of two host cell range variants of feline immunodeficiency virus  
JOURNAL J. Virol. 64, 4605-4613 (1990)  
STANDARD full staff\_review  
COMMENT Kindly provided in computer readable form by T. Phillips, Scripps Research Foundation, La Jolla CA.

The PPR isolate is approximately 91% homologous to the previously characterized Petaluma isolate. The latter infected feline kidney cells and the G355-5 cell line but replicated less efficiently on feline PBL's. In contrast, PPR productively infects PBL's but not the other cell lines. The authors point out interesting differences in the LTRs and coding regions.

The small orfs include: orf 1, in size and position similar to vif but without statistically significant homology; orf 2 that is "tat" like; orfs 3 (denoted D in [1]) and 4 (denoted H in [1]) seen also in the Petaluma strain.

FEATURES  
CDS  
Location/Qualifiers  
628..1980  
/note="gag polyprotein"  
/codon\_start=628  
<1869..5243  
/partial  
/note="pol polyprotein (NH2-terminus uncertain; AA at 1869)"  
/codon\_start=1869  
5236..5991  
/note="orf 1"  
/codon\_start=5236  
5992..6225  
/note="orf 2"  
/codon\_start=5992  
6263..8827  
/note="env polyprotein"  
/codon\_start=6263  
6709..6915  
/note="orf 3 (orf D in [J. Virol. 64, 4605-4613 (1990)])"  
/codon\_start=6709  
8950..9165  
/note="orf 4 (orf H in [J. Virol. 64, 4605-4613 (1990)])"  
/codon\_start=8950  
3605 a 1366 c 2092 g 2405 t  
BASE COUNT 3605 a 1366 c 2092 g 2405 t  
ORIGIN 5' terminus of 5'LTR.

Query Match 2.3%; Score 15; DB 3; Length 9468;  
Best Local Similarity 100.0%; Pred. No. 7.01e-01;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 8549 caaaattacaaat 8563  
|||||  
QY 596 CAAAATTACAAAT 610

RESULT 9 SIYCPZ 9811 bp ss-RNA VRL 11-DEC-1989  
LOCUS Simian immunodeficiency virus from chimpanzee, complete proviral genome;  
DEFINITION  
ACCESSION X52154  
SOURCE Chimpanzee immunodeficiency virus (CIV) proviral DNA, clone 1a.  
REFERENCE 1 (bases 1 to 9811)  
AUTHORS Huet,T., Cheynier,R., Meyerhaus,A., Roelants,G. and Wain-Hobson,S.  
TITLE Genetic organization of a chimpanzee lentivirus related to HIV-1  
JOURNAL Nature 345, 356-359 (1990)  
STANDARD

COMMENT Sequence kindly supplied in computer-readable form by T. Huet, Pasteur Institute. Clone 1a is infectious.  
The CPZ genome is more closely related to HIV-1s than to any other HIV or SIV, but it is more divergent from prototypical HIV-1 than any other isolate, with possible exception of the partially characterized ANR70. CPZ is especially different with respect to the vpu gene product.

FEATURES from to/span description

pept 827 2953 gag polyprotein  
pept < 2074 5157 pol polyprotein (NH2 terminus uncertain; AA at 2074)  
pept 5102 5683 vif protein  
pept 5623 5913 tat, exon 2 (first expressed exon)  
pept 5894 6111 tat, exon 3 (AA at 8420)  
pept 8419 8503 rev, exon 2 (first expressed exon)  
pept 6033 6111 rev, exon 3 (AA at 8421)  
pept 8714 8714 vpu protein  
pept 6120 6386 env polyprotein  
pept 6292 8856 nef protein  
pept 8870 9487  
BASE COUNT 3351 a 1837 c 2400 g 2223 t  
ORIGIN 5' terminus of the LTR-03.

Query Match 2.3%; Score 15; DB 3; Length 9811;  
Best Local Similarity 100.0%; Pred. No. 7.01e-01;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 2483 aacaattgataatg 2497  
|||||  
QY 549 AACAAATTGTATATG 563

RESULT 10 HIVU08687 273 bp ss-RNA VRL 01-APR-1994  
LOCUS Human immunodeficiency virus type 1, sample 017 from Brazil,  
DEFINITION C2V3 of env cds.  
ACCESSION U08687  
SOURCE Human immunodeficiency virus type 1 (HIV-1), sample 017 from Brazil; derived from a primary isolate.  
REFERENCE 1 (bases 1 to 273)  
AUTHORS WHO Global Programme on AIDS.  
COMMENT This sample is part of a set of sequences generated through the WHO Global Programme on AIDS. The virus was derived from an asymptomatic individual, from Brazil, whose route of infection is thought to be due to homosexual contact. The blood sample was taken in 1992. This env sequence clusters with HIV-1 B subtype sequences. The full name of this sequence is HIV192BR017WHO.01\_1gCR. It was presented in alignments in an abbreviated form in the April 94 Human Retroviruses and AIDS Compendium update as B2BR017W.01\_1gCR.

FEATURES Location/Qualifiers

CDS  
129 a 41 c 45 g 58 t  
/note="env polyprotein (AA at 1)"  
BASE COUNT 811 bp downstream from the beginning of env cds  
ORIGIN

Query Match 2.1%; Score 14; DB 3; Length 273;  
Best Local Similarity 100.0%; Pred. No. 5.15e+00;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Db 242 aacaattgataat 255  
 |||||||  
 QY 549 AACATTGATAAT 562

RESULT 11  
 LOCUS HIV08774 345 bp ds-DNA VRL 01-APR-1994  
 DEFINITION Human immunodeficiency virus type 1, sample 017 from Brazil, C2V3 of env cds.  
 U08774  
 ACCESSION  
 SOURCE Human immunodeficiency virus type 1 (HIV-1), sample 017 from Brazil; derived from a primary isolate.  
 1 (bases 1 to 345)  
 REFERENCE WHO Global Programme on AIDS.  
 AUTHORS This sample is part of a set of sequences generated through the WHO Global Programme on AIDS. The virus was derived from an asymptomatic individual, from Brazil, whose route of infection is thought to be due to homosexual contact. The blood sample was taken in 1992. This env sequence clusters with HIV-1 B subtype sequences. The full name of this sequence is HIV92BR017MO.01d1sCD; it was presented in alignments in an abbreviated form in the April 94 Human Retroviruses and AIDS compendium update as B2BR017W.01d1sCD.  
 COMMENT Location/Qualifiers  
 CDS  
 <1..>345  
 /note="env polypeptide (AA at 1)"

BASE COUNT 148 a 56 c 61 g 80 t  
 ORIGIN 829 bp from the beginning of env cds

Query Match 2.18; Score 14; DB 3; Length 345;  
 Best Local Similarity 100.08; Pred. No. 5.15e+00;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 224 aacaattgataat 237  
 |||||||  
 QY 549 AACATTGATAAT 562

RESULT 12  
 LOCUS VLVR 1174 bp ss-RNA VRL 10-APR-1990  
 DEFINITION Vigna lentivirus, strain K1514 cDNA; partial LTR, partial pol and partial env regions (two regions).  
 M25409  
 ACCESSION Vigna lentivirus (isolate K1514) cDNA clone from vicia provirus  
 SOURCE lambda-vis-109 clone.  
 1 (bases 1 to 1174)  
 REFERENCE Mazarin, V., Gourdou, I., Querat, G., Sauze, N. and Vigne, R.  
 AUTHORS Genetic structure and function of an early transcript of vicia  
 TITLE virus  
 JOURNAL J Virol 62, 4813-4818 (1988)

STANDARD  
 COMMENT This transcript was derived from the Vigna K1514 viral strain: 364 nucleotides truncated at the 3' end during cDNA synthesis were reconstructed with a fragment of the vicia provirus lambda vis 109 clone before cloning. The sequence consists of four exons. Comparison with the 9.4 kb full genomic sequence indicates that the four exons result from splicing at positions 303-304; 349-350 (coordinate 4888); and 501-502 (coordinate 6097) in the LTR, pol and env sequences. Transient-expression assays performed in eucaryotic cells demonstrated that this cDNA clone has a trans-acting effect on transcription of the vicia virus genes. The putative proteins Vep and stem were produced in vitro; Vep appears to be a rev-like protein.

FEATURES  
 from to/span description  
 pept 360 863 Vep1 protein (alternate start at 405)  
 pept 472 879 env stem (short transmembrane ) protein  
 BASE COUNT 405 a 230 c 326 g 213 t  
 ORIGIN

Query Match 2.18; Score 14; DB 3; Length 1174;  
 Best Local Similarity 100.08; Pred. No. 5.15e+00;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 512 aagtaaacagacatc 525  
 |||||||  
 QY 412 AAGTAACAGACATC 425

RESULT 13  
 LOCUS HIV2D205 7817 bp ss-RNA UNA 31-OCT-1989  
 DEFINITION Human immunodeficiency virus type 2 (HIV-2), isolate D205; gag, pol, vif, vpx, vpr, partial env.  
 X16109  
 ACCESSION 1 (bases 1 to 7817)  
 REFERENCE Dietrich, U., Adamski, M., Kreutz, R., Seip, A., Kuhnel, H. and Rudsamen-Waigmann, H.  
 AUTHORS A highly divergent HIV-2-related isolate  
 JOURNAL Nature 342, 948-950 (1990)  
 STANDARD  
 COMMENT This isolate, also called HIV2-ALT by the authors, was taken from a healthy Ghanaian woman. Tree analysis reveals that it likely arose prior to the branching of HIV2s and SIVs, i.e. the sooty mangabey and macaque isolates. [1] calls attention to an extension of the gag-pol overlap region resulting in a dilated (by 20 residues) pol ORF. The env cds begins at 6671; tat-exon1 at 6375 and rev-exon1 at 6601, with the possible donor splice site at 6665.

Kindly provided in computer readable form by H. Rudsamen-Waigmann, Georg Speyer Haus, Frankfurt.

FEATURES  
 CDS  
 <2299..5475  
 Location/Qualifiers  
 /note="pol polypeptide (NH2 terminus uncertain)"  
 /codon\_start=1  
 /translation="KTGLMKNRTHVKKPRKRTGGFFVRLTGKASQPHDPDSAS  
 GSDPTCPDEPSRSGDPTICAPCSGSSGDAKELADDETEREPREPTKINGIGGFG  
 EAPOFSLMRPVYKACIEGVSVIDTQVDSIVAGIELSGYKTYKINGIGGFG  
 TKRYKDEIEVGRKVRATITGDTPTINIRNRLNLGMLNTPVAKKYEKELP  
 GKDPKIRQWPLSKELIALKEICEKMEGQLEAPNTYNTPTAIRKKDKNRK  
 MLIDPRLINKVQDFTENVNWFPTROVAEKRRITVIDGAYFSPIDPNRQYTAFT  
 LPSVNAEPGRKRYIKVLPQMGKSSQICQYSKRVLDPEFRANSQVYIIQYMDIL  
 ASDRDLHDHNVFQLEKLELDMDGFEPEEFKDPDPFKMGLKMLKLOKLOL  
 EKEMVYNAIOKLVGNMAQLPGRITRICKLINSKMLITTEVQMTLAEALD  
 NKILDEQBSYIKERVPLEATVQKNAQMTYKIHQGNVLKGVYAKKNTING  
 VRLAHVQKIGKALYIWEIIPVEHLPVEEEDQMDWTQVTVLPEMDVSTPL  
 IRLAYNVKQPLEGRETYYDSCNRSKEKAGYVDGDKYKVEQTTQQADEL  
 AFALINDSEPOVNIIVDSQYVWGIIAOPETSPYAKIIEIMKEAVYGVNVA  
 HKIGGQNDVHIVSOGIROYLFEKTEPQAEHEKHYGNKELVHKGIPOLYAKOI  
 VNSDCKCOCKEALHGOVNDLGTWQDCHLEKTIIVAVHASGTEAVTQETG  
 ROTALFLKLASRPITLHLDNANFTSPSKVAVAWVGIQTFEYVYNQSGVE  
 AMNHKLKNOIDRLDQAVSIETVIMATHCNFRKGIGDMTPAEHLVNNITTEOEI  
 QEPQAKMLKQNFQVYREGDQMLKPGELMGEGAVIIVKGTETIKVPRRAKII  
 RHVGGKGLDQSDMEDTRQAREMAQSD"  
 1079..2644  
 /note="gag polypeptide"  
 /codon\_start=1  
 /translation="MGARGSVLSGKRTDELEKVLRLRGKKRWLKHVAVNDELDR  
 GLASLSEKSGCKILKVLAVLPVTESENLSLFTNYCVFLCHAEKVAVDTEAK  
 IAGHLLADTEKMPATKPPAPRGVNVPOQLAGNVFLSPRTNANAKYLEEK  
 FGAHVAGFOALSCTGCPYDINOMLVGGAAMQVLIIRINEADMDQOHPSP  
 MPAGQLDPRGSDIAGTSTYEDITQMYRQNVPPVGNITRRITQIGLOKCYMYNP  
 TNLIDIKQPREPQSTVVDREYKSLAEQDPAVKNMTOTLLIQANPCKLVTLK  
 GSNPTLEMLACGIGPGQKALMAELKALTPAIPFAAQAQKRGVTCN  
 CGKQGHAPROCRAPROGCMCGKGTGIMSCEPROAGFLGCMGKPRFPTQV  
 OGVPSPAPMNPAGMTPRATPAPPADPAVELKSYMQGRQGRSREBPYEVE  
 DLHLSLFGEDQ"  
 5396..6046  
 /note="vif protein"  
 /codon\_start=1  
 /translation="MEEKDMIVVPIRIPGRLEMRHSLIKLYKRTGELQOQVYVH  
 HVGAMWTCGRITFLPKGAMLEVGWVNTLTPRGFLSSAVRLTWYERFYTDP  
 KDADVLGHSYSCFSANVEYRAIRGKRIISYCNYPBAHSGOVPSLOFLAIRVOE  
 NGSGESATRKQRNRNRSTIRLARKNNNRQAQSGQFPAPRTTFPCLAEVLGTLA"  
 5877..6212







LOCUS HIVJRCSF 9540 bp ss-RNA VRL 01-DEC-1988  
 DEFINITION Human immunodeficiency virus type 1, isolate JRCSF; complete genome.  
 ACCESSION M38429  
 SOURCE HIV-1 proviral DNA from extracellular virus taken from cerebral spinal fluid (1986). Infectious clone.  
 REFERENCE 1 (bases 1 to 9540)  
 AUTHORS Koyanagi, S., Chen, I.S.-Y. et al.  
 JOURNAL Unpublished (1988).  
 STANDARD

Kindly provided in computer-readable form by Irvlin Chen, UCLA School of Medicine, Los Angeles. JRCSF and JRFL (see <HIVJRCFL> were isolated from cerebral spinal fluid and brain tissue of the patient JR, who died with Kaposi's sarcoma and severe AIDS encephalopathy (Science 236, 819-822, 1987). Both clones are infectious, but JRFL productively infects macrophages while JRCSF does not. (Peripheral blood was not available from the patient).

The JRCSF and JRFL env nucleotide sequences differ by at least 38; further characterization of them is forthcoming (Peng, S. et al., Nature 1990, in press). Both manifest insertions in nef previously reported for HIVBRVA.

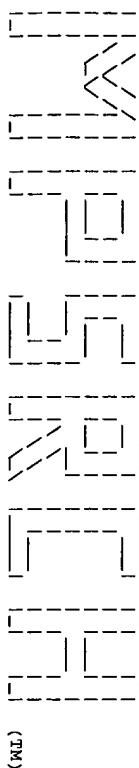
FEATURES	from to/span	description
pept	790	gag polyprotein
pept	< 2085	pol polyprotein (NH2-terminus uncertain; AA at 2085)
pept	5053	vif protein
pept	5571	vpr protein
pept	5842	tat protein, exon 2 (first expressed exon)
pept	8366	tat protein, exon 3 (AA at 8367)
pept	5981	rev protein, exon 2 (first expressed exon)
pept	8366	rev protein, exon 3 (AA at 8368)
pept	6073	vpu protein
pept	6236	env polyprotein
pept	8784	nef protein
LTR	1	5' LTR
LTR	9103	3' LTR
LTR	455	R repeat 5' copy
binding	378	SP1 binding site III
binding	389	SP1 binding site II
binding	400	SP1 binding site I
binding	637	primer (Lys-tRNA) binding site
BASE COUNT	3425 a	1691 c
ORIGIN	5'	terminus of 5' LTR.

Query Match 2.1%: Score 14; DB 3; Length 9540;  
 Best Local Similarity 100.0%: Pred. No. 5.15e+00;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 6196 ttaattgataaat 6209  
 |||  
 QY 491 TTAATTGATTAAT 504

Search completed: Sat Nov 27 15:04:49 1999  
 Job time : 12 secs.





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 15:00:41 1999; MasPar time 148.05 Seconds

Tabular output not generated. 955.232 Million cell updates/sec

Title: >US-09-103-287-3  
Description: (1.660) From US09103287.seq  
Perfect Score: 660  
N.A. Sequence: 1 ATTAAAGATTCGATGACA.....GCGTTTAAATGCTTAA 660  
Comp: TAAATTCCTAAGCCCTACTGT.....CGAATAATATACAAATAT

Scoring table: TABLE jmetric  
Gap 60

Mmatch STD : Dbase 0; Query 0

Searched: 271905 segs, 107135622 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

n-geneseg35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39 40:part40 41:part41 42:part42 43:part43  
44:part44 45:part45 46:part46 47:part47 48:part48  
49:part49 50:part50 51:part51 52:part52 53:part53  
54:part54 55:part55 56:part56 57:part57 58:part58  
59:part59 60:part60

Statistics: Mean 7.237; Variance 3.237; scale 2.236

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	505	76.5	2424	58	Staphylococcus aureus	0.00e+00
2	383	58.0	619	48	DNA encoding a Staphy	0.00e+00
3	381	57.7	660	60	Partial nucleotide se	0.00e+00
4	381	57.7	1351	60	UDP-N-acetylglucosamine	0.00e+00
5	20	3.0	2883	60	Enterococcus faecalis	1.63e-01
6	19	2.9	19	60	Murc polynucleotides	7.17e-01
7	19	2.9	1267	45	Streptococcus pneumonia	7.17e-01
8	19	2.9	1825	47	Streptococcus pneumonia	7.17e-01
9	19	2.9	11864	47	Streptococcus pneumonia	7.17e-01

10	18	2.7	154	58	Staphylococcus aureus	3.04e+00
11	18	2.7	58407	48	Methanococcus jannasch	3.04e+00
12	17	2.6	67	6	V11, Ox VL 5'-end PCR	1.23e+01
13	17	2.6	474	8	Human brain Expressed	1.23e+01
14	17	2.6	870	6	Ox VH-hinge-VL insert	1.23e+01
15	17	2.6	1370	2	Sequence of P-2 gene	1.23e+01
16	17	2.6	1491	12	Comamonas testosteron	1.23e+01
17	17	2.6	1608	8	Tapeal specific prom	1.23e+01
18	17	2.6	1890	11	DRG 206 genomic clon	1.23e+01
19	17	2.6	2269	12	Cytia gene which enc	1.23e+01
20	17	2.6	2546	2	Fragment pHS 53 homol	1.23e+01
21	17	2.6	3122	58	Staphylococcus aureus	1.23e+01
22	17	2.6	3359	52	Soybean seed coat per	1.23e+01
23	17	2.6	3628	38	Candida CARM2 gene.	1.23e+01
24	17	2.6	3628	38	Candida CARM2 gene.	1.23e+01
25	17	2.6	4410	15	Yeast MS11 gene.	1.23e+01
26	17	2.6	4700	47	Soybean seed coat per	1.23e+01
27	17	2.6	5873	10	Arabidopsis thaliana	1.23e+01
28	17	2.6	5880	10	Arabidopsis thaliana	1.23e+01
29	17	2.6	5890	10	Arabidopsis thaliana	1.23e+01
30	17	2.6	5890	10	Arabidopsis thaliana	1.23e+01
31	17	2.6	6211	47	Streptococcus pneumonia	1.23e+01
32	17	2.6	6295	30	Constitutive triple r	1.23e+01
33	17	2.6	6295	30	Constitutive triple r	1.23e+01
34	17	2.6	6312	15	Arabidopsis CTR1-2 mu	1.23e+01
35	17	2.6	6312	30	Constitutive triple r	1.23e+01
36	17	2.6	6312	30	Constitutive triple r	1.23e+01
37	17	2.6	6312	30	Constitutive triple r	1.23e+01
38	17	2.6	6312	15	Arabidopsis CTR1 geno	1.23e+01
39	17	2.6	6312	15	Arabidopsis CTR1-3 mu	1.23e+01
40	17	2.6	6312	15	Arabidopsis CTR1-1 mu	1.23e+01
41	17	2.6	6312	15	Arabidopsis CTR1-4 mu	1.23e+01
42	17	2.6	16484	60	Enterococcus faecalis	1.23e+01
43	16	2.4	6395	60	cDNA encoding mouse a	4.77e+01
44	16	2.4	28690	60	Enterococcus faecalis	4.77e+01

#### ALIGNMENTS

RESULT 1  
ID V74703 standard; DNA; 2424 BP.

AC V74703; 16-MAR-1999 (first entry)  
DE Staphylococcus aureus contig SEQ ID #392.  
KW Computer readable medium; vaccine; S. aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome; ds.  
OS Staphylococcus aureus.  
FH Key Location/Qualifiers  
FT misc\_feature 1141..1200  
FT /note="these bases represent a line of missing text in the sequence listing in the nucleotide numbering are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

EP-786519-A2.  
PF 07-JAN-1997; 100117.  
PR 05-JAN-1996; US-009861.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA, Rosen CA;  
PI WPI; 97-374922/35.  
DR Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
PT stored on computer readable medium and used in the production of  
PT anti-S. aureus vaccines  
PS Claim 1, Page 1287-1288; 3271pp; English.  
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
CC of the invention. The DNA sequences are recorded on a computer readable  
CC medium, preferably selected from a floppy or hard disk, random access  
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
CC the S. aureus DNA sequences allows putative functions to be assigned so















AS	detection: pneumonia; otitis media; meningitis; ss.
KW	Streptococcus pneumoniae.
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	2..1267
FT	/tag= a
FT	/product= "sp0070"
FT	/note= "no stop codon given"
PN	M09818930-A2.
PD	07-MAY-1998.
PF	30-OCT-1997; U19422.
PR	31-OCT-1996; US-029960.
PA	(HUMA-) HUMAN GENOME SCI INC.
P1	Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
DR	Wpi: 98-272224/24.
DR	P-PSDB: W55120.
PT	pneumoniae encoding antigenic peptide(s) from Streptococcus
PT	pneumoniae - or their epitope-containing fragments, useful in
PT	protective or therapeutic vaccines, and for diagnosis
PS	Claim 1; Page 72; 18pp; English.
CC	The present sequence encodes a protein from Streptococcus pneumoniae.
CC	The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC	can be useful in vaccines for inducing protective antibodies against
CC	Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC	pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC	are used to detect Streptococcus infection (by usual hybridisation or
CC	amplification methods), also for isolating Streptococcus genes or their
CC	allelic variants. The protein can be used similarly to detect specific
CC	antibodies in standard immunoassays, especially for diagnosing or
CC	monitoring infections. Antibodies which bind the protein are used to
CC	detect corresponding antigens, to purify the protein and for passive
CC	immunisation (optionally coupled to a toxin). Vaccines are administered
CC	e.g. by injection, orally or through the skin, typically at 0.01-1000
CC	(especially 10-300) mu g/ml per dose.
SQ	Sequence 1267 BP; 365 A; 261 G; 372 T;
Query Match	2.9%; Score 19; DB 46; Length 1267;
Best Local Similarity	100.0%; Pred. No. 7,17e-01;
Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Db	824 ggtgtaaacgcgttcata 842 
Oy	215 GGTTAAACGTCGGTTC A 233
RESULT	8
ID	V43027 standard; DNA; 1825 BP.
AC	V43027;
DT	09-NOV-1998 (first entry)
DE	Streptococcus pneumoniae polypeptide coding region.
RW	Polypeptide; ORF: open reading frame; infectious; bacterial;
KW	streptococcal; bacteremia; diagnosis; prophylaxis; ds.
OS	Streptococcus pneumoniae.
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	complement (256..423)
FT	/tag= a
FT	/note= "polypeptide"
FT	complement (731..868)
FT	/tag= b
FT	/note= "polypeptide"
PN	M09823631-A1.
PD	04-JUN-1998.
PF	24-NOV-1997; U21976.
PR	27-NOV-1996; US-031879.
PA	(SMIK ) SMITHKLINE BEECHAM CORP.
PA	(SWIK ) SMITHKLINE BEECHAM PLC.
P1	Black MT, Hodgson JE, Knowles DUC, Lonetto MA, Nicholas RO,
P1	Reid RH, Zarfos PN;
DR	WPI: 98-322654/28.
DR	P-PSDB: W62754, W63755.
PT	Streptococcus pneumoniae polynucleotides - useful for developing
PT	products for diagnosis, prevention and treatment of infections e.g.
PT	pneumonia, bacteraemia, meningitis or endocarditis
PS	Claim 1; Page 151-152; 181pp; English.

Query Match	2.9%	Score 19:	DB 47:	Length 1825:
Best Local Similarity 100.0%;		Pred. No. 7.17e-01:		
Matches 19; Conservative 0;		Mismatches 0;	Indels 0;	Gaps 0;
<p>Db 1160 tgaacgacgtttaacacc 1178         Cp 233 TGAACGACGTTTAACACC 215</p>				
<p>RESULT 9  ID VS2194 standard; DNA; 11864 BP.  AC VS2194;  DT 23-OCT-1998 (first entry)  DE Streptococcus pneumoniae genome fragment SEQ ID NO:61.  KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  KW computer readable medium; vaccine; pharmaceutical composition; ds.  OS Streptococcus pneumoniae.  PN M09818931-A2.  PD 07-MAY-1998.  PF 30-OCT-1997; U19588.  PR 31-OCT-1996; US-029960.  PA (HUMA-) HUMAN GENOME SCI INC.  PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,  PI Kunsch CA, Rosen CA;  PI WPI: 98-272225/24.  PT Computer-readable medium with recorded Streptococcus pneumoniae  PT polynucleotide sequences - useful in diagnostic kits and assays, and  PT pharmaceutical compositions and vaccines for Streptococcus  PT pneumoniae  PS Claim 1: Page 524-530; 1409pp; English.  CC The present invention describes a computer readable medium which has  CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  CC on it, or a representative fragment or a sequence at least 95% identical  CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  CC pneumoniae. The present invention also describes an isolated nucleic acid  CC molecule encoding a homologue of any of the fragments of the S.pneumoniae  CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  CC by a process comprising: (a) screening a genomic DNA library using as a  CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  CC to 391, identifying members of the library which contain sequences  CC that hybridize to the target sequence and isolating the nucleic acid  CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  CC from an organism, amplifying nucleic acid molecules whose nucleotide  CC sequence is homologous to amplification primers derived from the  CC fragment of the S. pneumoniae genome to prime the amplification and  CC isolating the amplified sequences. The computer readable medium can be  CC used in a computer-based system for identifying fragments of the  CC S. pneumoniae genome of commercial importance, or expression modulating  CC fragments of the S. pneumoniae genome. Products from the present  CC invention can be used in diagnosis kits and assays, and pharmaceutical  CC compositions and vaccines for S. pneumoniae.  CC Sequence 11864 BP; 3440 A; 2668 C; 2183 G; 3573 T;</p>				
Query Match	2.9%	Score 19:	DB 47:	Length 11864:
Best Local Similarity 100.0%;		Pred. No. 7.17e-01:		
Matches 19; Conservative 0;		Mismatches 0;	Indels 0;	Gaps 0;
<p>Db 855 tgaacgacgtttaacacc 873         Cp 233 TGAACGACGTTTAACACC 215</p>				
<p>RESULT 10  ID V75600 standard; DNA; 154 BP.</p>				



AC V75600; (first entry)  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SPQ ID #1289.  
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 OS Staphylococcus aureus.  
 PD 26-FEB-1998.  
 PN BP-786519-A2.  
 PD 30-JUL-1997.  
 PF 07-JAN-1997: 100117.  
 PR 05-JAN-1996; US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
 PI Rosen CA;  
 DR MPI: 97-374922/35.  
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
 stored on computer readable medium and used in the production of  
 anti-S.aureus vaccines  
 PT Claim 1: Page 1907: 3271pp; English.  
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 of the invention. The DNA sequences are recorded on a computer readable  
 medium, preferably selected from a floppy or hard disk, random access  
 memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 the S.aureus DNA sequences allows putative functions to be assigned so  
 that protein-encoding or regulatory regions of commercial, therapeutic or  
 industrial importance can be obtained. Specifically, sequences which are  
 likely to encode antigens have been identified and these polypeptides can  
 be used in a vaccine composition against S.aureus infection. The  
 polypeptides can also be used in a kit for the immunodetection of  
 S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 skin and surgical wound infections, scalded skin syndrome, toxic shock  
 syndrome, etc. Organisms transformed with the DNA sequences can be used  
 for recombinant production of the polypeptides. The new DNA sequences  
 (and their fragments) are useful as primers or probes for isolating  
 homologues of any of the S.aureus DNA sequences contained on the  
 computer readable medium.  
 CC Sequence 154 BP: 42 A; 32 C; 25 G; 54 T;  
 SQ

Query Match 2.7%; Score 18; DB 58; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 3.04e+00;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 23 gccgttaataatgtttat 40  
 ||||||||||||||||  
 OY 641 GCCTTTAATATCTTAT 658

RESULT 11  
 ID V21210 standard; DNA; 58407 BP.  
 AC V21210;  
 DT 10-NOV-1998 (first entry)  
 DE Methanococcus jannaschii large circular extrachromosomal element.  
 KW Methanococcus jannaschii; methanogenic archaeon; circular chromosome;  
 OS genome; autotrophic; extrachromosomal element; identification; ds.  
 PD 26-FEB-1998.  
 PN WO9807830-A2.  
 PD 26-FEB-1998.  
 PF 22-AUG-1997: U14900.  
 PR 22-AUG-1996; US-024428.  
 PA (GENO-) INST GENOMIC RES.  
 PA (UNITI ) UNIV ILLINOIS RES.  
 PA (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
 PI Bult CJ, Smith HO, Venter JC, White OR, Woese CR;  
 DR MPI: 98-169145/15.  
 CC Complete genome sequence of methano-genic archaeon, Methanococcus  
 jannaschii - useful in identification of M. jannaschii genome  
 fragment  
 PT Claim 13: Page 585-600; 614pp; English.  
 CC The present sequence represents the large circular extrachromosomal  
 element sequence of the Methanococcus jannaschii circular chromosome. The  
 present invention describes M. jannaschii open reading frames from the

CC genome sequence. The invention also describes a computer based system  
 for identifying fragments of the M. jannaschii genome that are  
 homologous to target nucleotide sequences, comprising: (a) data storage  
 CC means comprising the nucleotide sequence of the 1664976, 58407 or 16550  
 CC bp sequence (see V21209, V21210 and V21211), or a nucleotide sequence at  
 CC least 99.98 identical to it; (b) search means for comparing a target  
 CC sequence to the nucleotide sequence of the data storage means to  
 CC identify a homologous sequence, and (c) retrieval means for obtaining  
 CC the homologous sequence. The method, which is based on whole genome  
 CC random sequencing of an autotrophic archaeon M. jannaschii, the genome  
 CC of which consists of 3 physically distinct elements, a large circular  
 CC chromosome (the 1664976 bp sequence given in V21209), a large circular  
 CC extra-chromosomal element (the 58407 bp sequence given in V21210), and a  
 CC small circular extra-chromosomal element (the 16550 bp sequence given in  
 CC V21211), can be used in the identification of M. jannaschii genome  
 fragment.  
 CC Sequence 58407 BP: 20163 A; 9034 C; 7447 G; 21761 T;  
 SQ

Query Match 2.7%; Score 18; DB 48; Length 58407;  
 Best Local Similarity 100.0%; Pred. No. 3.04e+00;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 56456 tttaataatgaatcct 56473  
 ||||||||||||||||  
 CP 503 TTTATCAATTAACTT 486

RESULT 12  
 ID Q36987 standard; DNA; 67 BP.  
 AC Q36987;  
 DT 08-JUN-1993 (first entry)  
 DE VLI, Ox VL 5'-end PCR primer.  
 KW Spacer peptide; secretible; single chain; fusion protein; antibody;  
 KW scab: recombinant; polymerase chain reaction; ss.  
 OS Synthetic.  
 PN PI9103434-A.  
 PN WO9302198-A. (First Major Country Equivalent).  
 PD 17-JAN-1992.  
 PF 16-JUL-1991: 913434.  
 PF 16-JUL-1990; US-552751.  
 PA (TERE-) TECH RES CENT FINLAND.  
 PI Alifhan K, Knowles JKC, Laakkonen ML, Stizmann D, Takkinen K, Teeri TT;  
 DR MPI: 92-134225/17.  
 CC Prod. of single chain fusion protein, pref. antibody - comprises  
 PT transforming host cells, e.g. E. coli with expression constructs  
 PT composed of proteins or domains, linked by spacer peptide(s)  
 PS Example: Page 28; 56pp; English.  
 CC The sequence is that of the PCR primer VLI, which is used in the  
 CC modification, for in-frame fusions with the hinge coding region,  
 CC of the 5'-end of Ox VL cDNA.  
 CC Sequence 67 BP: 16 A; 22 C; 11 G; 18 T;  
 SQ

Query Match 2.6%; Score 17; DB 6; Length 67;  
 Best Local Similarity 100.0%; Pred. No. 1.23e+01;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 37 cattatcaaatgttct 53  
 ||||||||||||||||  
 CP 563 CATTATCAAAATGTCT 547

RESULT 13  
 ID Q59532 standard; cDNA; 474 BP.  
 AC Q59532;  
 DT 16-MAR-1994 (first entry)  
 DE Human Brain Expressed Sequence Tag EST00527.  
 KW Gene transcription product; genetic markers; tagging; in vivo;  
 KW transcription; mapping; locations; chromosomes; chromosomal; ss.  
 OS Homo sapiens.  
 PN WO9316178-A.  
 PD 19-AUG-1993.  
 PF 12-FEB-1993; U01294.  
 PF 12-FEB-1992; US-837195.



PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 PI Adams MD, Moreno RF, Venter CJ;  
 DR WPI: 93-272882/34.  
 PT Enriched oligonucleotides and corresp. sequences - used as  
 PT markers for human genes transcribed in-vivo, facilitate tagging  
 PT of most human genes  
 PS Example 4: Page 190; 500pp; English.  
 CC The Expressed Sequence tag was isolated from a human brain cDNA  
 CC library as part of a large set of ESTs which can be used as markers  
 CC for human genes transcribed in vivo. They can be used to facilitate  
 CC tagging of most human genes, for mapping locations of expressed genes  
 CC on chromosomes, for individual or forensic identification, for mapping  
 CC locations of disease-associated genes, for identification of tissue  
 CC type, and for prepn. of antisense sequences, probes and constructs.  
 CC EST00527 has a "poor" coding probability as evaluated using the  
 CC coding-region prediction program CRM. See also Q59041-Q61440.  
 SQ Sequence 474 BP; 155 A; 88 C; 75 G; 153 T;  
 Query Match 2.6%; Score 17; DB 8; Length 474;  
 Best Local Similarity 100.0%; Pred. No. 1.23e+01;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 372 catttcaatgaattt 388  
 QY 384 CATTTTAAATGAATTT 400  
 RESULT 14  
 ID Q36982 standard: cDNA: 870 BP.  
 AC Q36982;  
 DT 08-JUN-1993 (first entry)  
 DE Ox VH-hinge-VL insert.  
 KM Spacer peptide; secreted; single chain; fusion protein; antibody;  
 KM scAb; recombinant; rDNA; linker; Bos taurus; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_feature 61..126  
 FT /\*tag- a  
 FT /\*note- "SS"  
 FT misc\_feature 127..465  
 FT /\*tag- b  
 FT /\*note- "Ox VH"  
 FT misc\_feature 466..549  
 FT /\*tag- c  
 FT /\*note- "CBHI hinge"  
 FT misc\_feature 550..870  
 FT /\*tag- d  
 FT /\*note- "VL"  
 FT FT  
 PN FT9103434-A. (First Major Country Equivalent).  
 PN W09302198-A.  
 PD 17-JAN-1992.  
 PF 16-JUL-1991; 913434.  
 PR 16-JUL-1990; US-552751.  
 PA (TERR-) TECH RES CENT FINLAND.  
 PI Alifthan K, Knowles JKC, Laukkanen ML, Sizmann D, Takkinen K, Teeri TT;  
 DR WPI: 92-134225/17.  
 DR P-PSDB: R32569.  
 PT Prod. of single chain fusion protein, pref. antibody - comprises  
 PT transforming host cells, e.g. E. coli with expression constructs  
 PT composed of proteins or domains, linked by spacer peptide(s)  
 PS Example; Fig 4; 56pp; English.  
 CC The sequence is that of an Ox VH-CBHI hinge-VL insert which was  
 CC used as part of a method for cloning secreted, biologically  
 CC active single chain antibodies (scAbs) and other secreted fusion  
 CC proteins having at least 2 distinct functional proteins or domains.  
 SQ Sequence 870 BP; 213 A; 239 C; 222 G; 196 T;  
 Query Match 2.6%; Score 17; DB 6; Length 870;  
 Best Local Similarity 100.0%; Pred. No. 1.23e+01;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 544 cattatcaattgtctt 560  
 ||||||||||||||||||

CP 563 CATTATCAATGTCTCT 547  
 RESULT 15  
 ID N80591 standard: DNA: 1370 BP.  
 AC N80591;  
 DT 12-JAN-1991 (first entry)  
 DE Sequence of P-2 gene of *Bacillus thuringiensis* var. *kurstaki*  
 DE insecticide; lepidoptera; diptera; P-2 delta-endotoxin;  
 KM parasporal crystal protein; ss.  
 OS *Bacillus thuringiensis*.  
 FH Key Location/Qualifiers  
 FT cds 156..1928  
 FT /\*tag- a  
 PN W08808034-A.  
 PD 20-OCT-1988.  
 PF 07-APR-1988; U01132.  
 PR 16-APR-1987; US-039542.  
 PA (ECCG-) Ecogen Inc.  
 PI Donovan WP;  
 DR WPI: 88-307569/43.  
 DR P-PSDB: P80548.  
 PT *Bacillus thuringiensis* P-2 toxic gene and expressed protein -  
 PT used in insecticidal compns. having activity against  
 PT lepidoptera and diptera insects  
 PS Claim 1; Fig 2(1) and Fig 2(3); 66pp; English.  
 CC A gene for *Bacillus thuringiensis* (B.t.) P-2 toxin having the DNA sequence  
 CC specified in Fig 2 (N80591), or any portion or deriv. of it, is claimed.  
 CC Also claimed is the protein encoded by the gene, having the corresponding  
 CC specified AA sequence in P80548. In isolating the P-2 gene, the P-2  
 CC protein was purified from a donor strain of B.t. var. *kurstaki* and the  
 CC partial AA sequence of the P-2 protein was determined. P-2 gene-specific  
 CC oligo probe (N80590) was synthesised based on the AA sequence. The P-2  
 CC toxin encoded by the cloned gene has insecticidal activity against  
 CC lepidoptera and diptera insects.  
 CC N.B. The DNA/protein sequence in N80591/P80548 is as given in the  
 CC specification and is missing bases 901-1800 and the corresp. AAs.  
 SQ Sequence 1370 BP; 486 A; 185 C; 236 G; 463 T;  
 Query Match 2.6%; Score 17; DB 2; Length 1370;  
 Best Local Similarity 100.0%; Pred. No. 1.23e+01;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 9 acaagaattatgata 25  
 ||||||||||||||||||  
 QY 484 ACAAGATTATGATA 500

Search completed: Sat Nov 27 15:04:20 1999  
 Job time : 219 secs.



**THIS PAGE BLANK (USPTO)**







Query Match 2.6%; Score 17; DB 1; Length 1608;  
Best Local Similarity 100.0%; Pred. No. 8.93e-01;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1448 TGATTAATGAGGTG 1464  
|||||  
QY 496 TGATTAATGAGGTG 512

RESULT 2  
ID 5312912-3 STANDARD; DNA; UNC; 2047 BP.  
AC xxxxxx  
DE 01-JAN-1900  
CC Patent No. 5312912.  
CC Patent No. 5312912  
CC APPLICANT: HADWIGER, LEE A.; CHIANG, CHIN C.; HOROVITZ,  
CC DANIEL A.  
CC TITLE OF INVENTION: PROCEDURES AND REGULATORY DNA SEQUENCES  
CC FOR GENETICALLY ENGINEERING DISEASE RESISTANCE AND OTHER  
CC INDUCIBLE TRAITS IN PLANTS  
CC NUMBER OF SEQUENCES: 9  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/393,301  
CC FILING DATE: 13-JUN-1989  
CC SEQ ID NO:3  
CC LENGTH: 1890  
CC Sequence 2047 BP; 656 A; 270 C; 254 G; 704 T; 163 other;

Query Match 2.6%; Score 17; DB 5; Length 1690;  
Best Local Similarity 100.0%; Pred. No. 8.93e-01;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1487 AATGAGATCTATTAA 1503  
|||||  
QY 524 AATGAGATCTATTAA 540

RESULT 3  
ID PCT-US93-07347-4 STANDARD; DNA; UNC; 5873 BP.  
AC xxxxxx  
DE Sequence 4, Application PC/TUS9307347  
CC Sequence 4, Application PC/TUS9307347  
CC GENERAL INFORMATION:  
CC APPLICANT: Ecker, Joseph R.  
CC APPLICANT: Kieher, Joseph J.  
CC TITLE OF INVENTION: Constitutive Triple Response Gene and  
CC TITLE OF INVENTION: Mutations  
CC NUMBER OF SEQUENCES: 6  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz and  
CC ADDRESSEE: Norris  
CC STREET: One Liberty Place - 46th Floor  
CC CITY: Philadelphia  
CC STATE: PA  
CC COUNTRY: U.S.A.  
CC ZIP: 19103  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US93/07347  
CC FILING DATE: 19930805  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Miller, Suzanne E.  
CC REGISTRATION NUMBER: 32,279  
CC REFERENCE/DOCKET NUMBER: UPN-1086  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 215-568-3100  
CC INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
LENGTH: 5873 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
SQ SEQUENCE 5873 BP; 1664 A; 1052 C; 1238 G; 1919 T; 0 OTHER.

Query Match 2.6%; Score 17; DB 4; Length 5873;  
Best Local Similarity 100.0%; Pred. No. 8.93e-01;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2560 GCGTGTGATGTATGT 2576  
|||||  
QY 59 GCGTGTGATGTATGT 75

RESULT 4  
ID US-07-928-464-4 STANDARD; DNA; UNC; 5873 BP.  
AC xxxxxx  
DE Sequence 4, Application US/07928464  
CC Sequence 4, Application US/07928464  
CC Patent No. 5367065  
CC GENERAL INFORMATION:  
CC APPLICANT: Ecker, Joseph R.  
CC APPLICANT: Kieher, Joseph J.  
CC TITLE OF INVENTION: Constitutive Triple Response Gene and  
CC TITLE OF INVENTION: Mutations  
CC NUMBER OF SEQUENCES: 6  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz and  
CC ADDRESSEE: No. 5367065ris  
CC STREET: One Liberty Place - 46th Floor  
CC CITY: Philadelphia  
CC STATE: PA  
CC COUNTRY: U.S.A.  
CC ZIP: 19103  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: IBM PC compatible  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/928,464  
CC FILING DATE: 19920810  
CC CLASSIFICATION: 800  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Miller, Suzanne E.  
CC REGISTRATION NUMBER: 32,279  
CC REFERENCE/DOCKET NUMBER: UPN-1086  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 215-568-3100  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 5873 base pairs  
CC TYPE: NUCLEIC ACID  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
SQ SEQUENCE 5873 BP; 1664 A; 1052 C; 1238 G; 1919 T; 0 OTHER.

Query Match 2.6%; Score 17; DB 1; Length 5873;  
Best Local Similarity 100.0%; Pred. No. 8.93e-01;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2560 GCGTGTGATGTATGT 2576  
|||||  
QY 59 GCGTGTGATGTATGT 75

RESULT 5  
ID US-07-928-464-3 STANDARD; DNA; UNC; 5890 BP.



AC xxxxxx  
 DE Sequence 3, Application US/07928464  
 CC Sequence 3, Application US/07928464  
 CC Patent No. 5367065  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Ecker, Joseph R.  
 CC APPLICANT: Kieber, Joseph J.  
 CC TITLE OF INVENTION: Constitutive Triple Response Gene and  
 CC TITLE OF INVENTION: Mutations  
 CC NUMBER OF SEQUENCES: 6  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz and  
 CC ADDRESSEE: No. 5367065rls  
 CC STREET: One Liberty Place - 46th Floor  
 CC CITY: Philadelphia  
 CC STATE: PA  
 CC COUNTRY: U.S.A.  
 CC ZIP: 19103  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/07/928,464  
 CC FILING DATE: 19920810  
 CC CLASSIFICATION: 800  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Miller, Suzanne E.  
 CC REGISTRATION NUMBER: 32,279  
 CC REFERENCE/DOCKET NUMBER: UPN-1086  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: 215-568-3100  
 CC INFORMATION FOR SEQ. ID NO. 3:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 5890 base pairs  
 CC TYPE: NUCLEIC ACID  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: DNA (genomic)  
 CC FEATURE:  
 CC NAME/KEY: Intron  
 CC LOCATION: 1..353  
 CC FEATURE:  
 CC NAME/KEY: exon  
 CC LOCATION: 354..1001  
 CC FEATURE:  
 CC NAME/KEY: Intron  
 CC LOCATION: 1002..1176  
 CC FEATURE:  
 CC NAME/KEY: exon  
 CC LOCATION: 1177..1477  
 CC FEATURE:  
 CC NAME/KEY: Intron  
 CC LOCATION: 1478..1574  
 CC FEATURE:  
 CC NAME/KEY: exon  
 CC LOCATION: 1575..1719  
 CC FEATURE:  
 CC NAME/KEY: Intron  
 CC LOCATION: 1720..1936  
 CC FEATURE:  
 CC NAME/KEY: exon  
 CC LOCATION: 1937..2038  
 CC FEATURE:  
 CC NAME/KEY: Intron  
 CC LOCATION: 2039..2173  
 CC FEATURE:  
 CC NAME/KEY: exon  
 CC LOCATION: 2174..2379  
 CC FEATURE:  
 CC NAME/KEY: Intron

CC LOCATION: 2380..2736  
 CC FEATURE:  
 CC NAME/KEY: exon  
 CC LOCATION: 2737..3012  
 CC FEATURE:  
 CC NAME/KEY: Intron  
 CC LOCATION: 3013..3202  
 CC FEATURE:  
 CC NAME/KEY: exon  
 CC LOCATION: 3203..3243  
 CC FEATURE:  
 CC NAME/KEY: Intron  
 CC LOCATION: 3244..3519  
 CC FEATURE:  
 CC NAME/KEY: exon  
 CC LOCATION: 3520..3588  
 CC FEATURE:  
 CC NAME/KEY: Intron  
 CC LOCATION: 3589..3668  
 CC FEATURE:  
 CC NAME/KEY: exon  
 CC LOCATION: 3669..3769  
 CC FEATURE:  
 CC NAME/KEY: Intron  
 CC LOCATION: 3770..3858  
 CC FEATURE:  
 CC NAME/KEY: exon  
 CC LOCATION: 3859..3943  
 CC FEATURE:  
 CC NAME/KEY: Intron  
 CC LOCATION: 3944..4037  
 CC FEATURE:  
 CC NAME/KEY: exon  
 CC LOCATION: 4038..4136  
 CC FEATURE:  
 CC NAME/KEY: Intron  
 CC LOCATION: 4137..4369  
 CC FEATURE:  
 CC NAME/KEY: exon  
 CC LOCATION: 4370..4438  
 CC FEATURE:  
 CC NAME/KEY: Intron  
 CC LOCATION: 4439..4541  
 CC FEATURE:  
 CC NAME/KEY: exon  
 CC LOCATION: 4542..4673  
 CC FEATURE:  
 CC NAME/KEY: Intron  
 CC LOCATION: 4674..4787  
 CC FEATURE:  
 CC NAME/KEY: exon  
 CC LOCATION: 4788..4882  
 CC FEATURE:  
 CC NAME/KEY: Intron  
 CC LOCATION: 4883..4959  
 CC FEATURE:  
 CC NAME/KEY: exon  
 CC LOCATION: 4960..5056  
 CC FEATURE:  
 CC NAME/KEY: Intron  
 CC LOCATION: 5057..5890  
 CC SEQUENCE 5890 BP: 1666 A; 1055 C; 1240 G; 1929 T; 0 OTHER.  
 SQ  
 Query Match 2.58; Score 17; DB 1; Length 5890;  
 Best Local Similarity 100.0%; Pred. No. 8.93e-01;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 2577 GCGTGTGATGTGATGT 2593  
 QY 59 GCGTGTGATGTGATGT 75  
 RESULT 6



ID PCT-US93-07347-6 STANDARD; DNA; UNC; 5890 BP.  
AC xxxxxx  
DE Sequence 6, Application PC/TUS9307347  
CC Sequence 6, Application PC/TUS9307347  
CC GENERAL INFORMATION:  
CC APPLICANT: Ecker, Joseph R.  
CC APPLICANT: Kieber, Joseph J.  
CC TITLE OF INVENTION: Constitutive Triple Response Gene and  
CC TITLE OF INVENTION: Mutations  
CC NUMBER OF SEQUENCES: 6  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz and  
CC ADDRESSEE: Norris  
CC STREET: One Liberty Place - 46th Floor  
CC CITY: Philadelphia  
CC STATE: PA  
CC COUNTRY: U.S.A.  
CC ZIP: 19103  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US93/07347  
CC FILING DATE: 19930805  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Miller, Suzanne E.  
CC REGISTRATION NUMBER: 32,279  
CC REFERENCE/DOCKET NUMBER: UPN-1086  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 215-568-3100  
CC INFORMATION FOR SEQ ID NO: 6:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 5890 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC SEQUENCE 5890 BP; 1667 A; 1055 C; 1240 G; 1928 T; 0 OTHER.  
SQ  
Query Match 2.6%; Score 17; DB 4; Length 5890;  
Best Local Similarity 100.0%; Pred. No. 8.93e-01;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 2577 GCGTGTGATGTATGT 2593  
OY 59 GCGTGTGATGTATGT 75  
RESULT 7  
ID PCT-US93-07347-3 STANDARD; DNA; UNC; 5890 BP.  
AC xxxxxx  
DE Sequence 3, Application PC/TUS9307347  
CC Sequence 3, Application PC/TUS9307347  
CC GENERAL INFORMATION:  
CC APPLICANT: Ecker, Joseph R.  
CC APPLICANT: Kieber, Joseph J.  
CC TITLE OF INVENTION: Constitutive Triple Response Gene and  
CC TITLE OF INVENTION: Mutations  
CC NUMBER OF SEQUENCES: 6  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz and  
CC ADDRESSEE: Norris  
CC STREET: One Liberty Place - 46th Floor  
CC CITY: Philadelphia  
CC STATE: PA  
CC COUNTRY: U.S.A.  
CC ZIP: 19103  
CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US93/07347  
CC FILING DATE: 19930805  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Miller, Suzanne E.  
CC REGISTRATION NUMBER: 32,279  
CC REFERENCE/DOCKET NUMBER: UPN-1086  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 215-568-3100  
CC INFORMATION FOR SEQ ID NO: 3:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 5890 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC FEATURE:  
CC NAME/KEY: Intron  
CC LOCATION: 1..353  
CC FEATURE:  
CC NAME/KEY: exon  
CC LOCATION: 354..1001  
CC FEATURE:  
CC NAME/KEY: Intron  
CC LOCATION: 1002..1176  
CC FEATURE:  
CC NAME/KEY: exon  
CC LOCATION: 1177..1477  
CC FEATURE:  
CC NAME/KEY: Intron  
CC LOCATION: 1478..1574  
CC FEATURE:  
CC NAME/KEY: exon  
CC LOCATION: 1575..1719  
CC FEATURE:  
CC NAME/KEY: Intron  
CC LOCATION: 1720..1936  
CC FEATURE:  
CC NAME/KEY: exon  
CC LOCATION: 1937..2038  
CC FEATURE:  
CC NAME/KEY: Intron  
CC LOCATION: 2039..2173  
CC FEATURE:  
CC NAME/KEY: exon  
CC LOCATION: 2174..2379  
CC FEATURE:  
CC NAME/KEY: Intron  
CC LOCATION: 2380..2736  
CC FEATURE:  
CC NAME/KEY: exon  
CC LOCATION: 2737..3012  
CC FEATURE:  
CC NAME/KEY: Intron  
CC LOCATION: 3013..3202  
CC FEATURE:  
CC NAME/KEY: exon  
CC LOCATION: 3203..3243  
CC FEATURE:  
CC NAME/KEY: Intron  
CC LOCATION: 3244..3519  
CC FEATURE:  
CC NAME/KEY: exon  
CC LOCATION: 3520..3588  
CC FEATURE:  
CC NAME/KEY: Intron  
CC LOCATION: 3589..3668  
CC FEATURE:







CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
SQ SEQUENCE 5890 BP; 1666 A; 1054 C; 1240 G; 1930 T; 0 OTHER.  
Query Match 2.6%; Score 17; DB 1; Length 5890;  
Best Local Similarity 100.0%; Pred. No. 8.93e-01;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 2577 GCGTGTGATGTATGT 2593  
|||  
OY 59 GCGTGTGATGTATGT 75

RESULT 10  
ID PCT-US93-07347-5 STANDARD; DNA; UNC; 5890 BP.  
AC xxxxxx  
DE Sequence 5, Application PC/TUS9307347  
CC Sequence 5, Application PC/TUS9307347  
CC GENERAL INFORMATION:  
CC APPLICANT: Ecker, Joseph R.  
CC APPLICANT: Kieker, Joseph J.  
CC TITLE OF INVENTION: Constitutive Triple Response Gene and  
CC NUMBER OF SEQUENCES: 6  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz and  
CC ATTORNEY/AGENT INFORMATION:  
CC STREET: One Liberty Place - 46th Floor  
CC CITY: Philadelphia  
CC STATE: PA  
CC COUNTRY: U.S.A.  
CC ZIP: 19103  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US93/07347  
CC FILING DATE: 19930805  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Miller, Suzanne E.  
CC REGISTRATION NUMBER: 32,279  
CC REFERENCE/DOCKET NUMBER: UPN-1086  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 215-568-3100  
CC INFORMATION FOR SEQ ID NO: 5:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 5890 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
SQ SEQUENCE 5890 BP; 1666 A; 1054 C; 1240 G; 1930 T; 0 OTHER.  
Query Match 2.6%; Score 17; DB 4; Length 5890;  
Best Local Similarity 100.0%; Pred. No. 8.93e-01;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 2577 GCGTGTGATGTATGT 2593  
|||  
OY 59 GCGTGTGATGTATGT 75

CC Patent No. 5444166  
CC GENERAL INFORMATION:  
CC APPLICANT: Ecker, Joseph R.  
CC APPLICANT: Kieker, Joseph J.  
CC TITLE OF INVENTION: Constitutive Triple Response Gene  
CC NUMBER OF SEQUENCES: 13  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and  
CC ADDRESSEE: No. 5444166ris  
CC STREET: One Liberty Place - 46th Floor  
CC CITY: Philadelphia  
CC STATE: PA  
CC COUNTRY: U.S.A.  
CC ZIP: 19103  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Wordperfect 5.1  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/003,311B  
CC FILING DATE: January 12, 1993  
CC CLASSIFICATION: 800  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/928,464  
CC FILING DATE: August 10, 1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Lori Y. Beardsell  
CC REGISTRATION NUMBER: 34,293  
CC REFERENCE/DOCKET NUMBER: UPN-1108  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 215-568-3100  
CC TELEFAX: 215-568-3439  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 6295 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
SQ SEQUENCE 6295 BP; 1815 A; 1104 C; 1295 G; 2081 T; 0 OTHER.  
Query Match 2.6%; Score 17; DB 1; Length 6295;  
Best Local Similarity 100.0%; Pred. No. 8.93e-01;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 2982 GCGTGTGATGTATGT 2998  
|||  
OY 59 GCGTGTGATGTATGT 75

RESULT 12  
ID US-08-261-432-4 STANDARD; DNA; UNC; 6295 BP.  
AC xxxxxx  
DE Sequence 4, Application US/08261432  
CC Sequence 4, Application US/08261432  
CC Patent No. 5602322  
CC GENERAL INFORMATION:  
CC APPLICANT: Ecker, Joseph R.  
CC APPLICANT: Kieker, Joseph J.  
CC TITLE OF INVENTION: Constitutive Triple Response Gene  
CC NUMBER OF SEQUENCES: 13  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and  
CC ADDRESSEE: No. 5602322ris  
CC STREET: One Liberty Place - 46th Floor  
CC CITY: Philadelphia  
CC STATE: PA  
CC COUNTRY: U.S.A.  
CC ZIP: 19103



CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Wordperfect 5.1  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/261,432  
CC FILING DATE: June 17, 1994  
CC CLASSIFICATION: 800  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/003,311  
CC FILING DATE: January 12, 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Lori Y. Beardell  
CC REGISTRATION NUMBER: 34,293  
CC REFERENCE/DOCKET NUMBER: UPN-1864  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 215-568-3100  
CC TELEFAX: 215-568-3439  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 6295 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC SEQUENCE 6295 BP; 1815 A; 1104 C; 1295 G; 2081 T; 0 OTHER.  
Query Match 2.6%; Score 17; DB 1; Length 6295;  
Best Local Similarity 100.0%; Pred. No. 8.93e-01;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 2982 GCTGTGATGTGTGTGT 2998  
|||  
QY 59 GCTGTGATGTGTGTGT 75  
RESULT 13  
ID US-08-261-432-3 STANDARD; DNA; UNC; 6312 BP.  
AC xxxxxx  
DE Sequence 3, Application US/08261432  
CC Sequence 3, Application US/08261432  
CC Patent No. 5602322  
CC GENERAL INFORMATION:  
CC APPLICANT: Ecker, Joseph R.  
CC TITLE OF INVENTION: Constitutive Triple Response Gene  
CC TITLE OF INVENTION: and Mutations  
CC NUMBER OF SEQUENCES: 13  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and  
CC ADDRESSEE: No. 5602322r1s  
CC STREET: One Liberty Place - 46th Floor  
CC CITY: Philadelphia  
CC STATE: PA  
CC COUNTRY: U.S.A.  
CC ZIP: 19103  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Wordperfect 5.1  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/261,432  
CC FILING DATE: June 17, 1994  
CC CLASSIFICATION: 800  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/003,311  
CC FILING DATE: January 12, 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Lori Y. Beardell  
CC REGISTRATION NUMBER: 34,293

CC REFERENCE/DOCKET NUMBER: UPN-1864  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 215-568-3100  
CC TELEFAX: 215-568-3439  
CC INFORMATION FOR SEQ ID NO: 3:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 6312 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC SEQUENCE 6312 BP; 1817 A; 1107 C; 1297 G; 2091 T; 0 OTHER.  
Query Match 2.6%; Score 17; DB 1; Length 6312;  
Best Local Similarity 100.0%; Pred. No. 8.93e-01;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 2999 GCTGTGATGTGTGTGT 3015  
|||  
QY 59 GCTGTGATGTGTGTGT 75  
RESULT 14  
ID US-08-261-432-5 STANDARD; DNA; UNC; 6312 BP.  
AC xxxxxx  
DE Sequence 5, Application US/08261432  
CC Sequence 5, Application US/08261432  
CC Patent No. 5602322  
CC GENERAL INFORMATION:  
CC APPLICANT: Ecker, Joseph R.  
CC TITLE OF INVENTION: Constitutive Triple Response Gene  
CC TITLE OF INVENTION: and Mutations  
CC NUMBER OF SEQUENCES: 13  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and  
CC ADDRESSEE: No. 5602322r1s  
CC STREET: One Liberty Place - 46th Floor  
CC CITY: Philadelphia  
CC STATE: PA  
CC COUNTRY: U.S.A.  
CC ZIP: 19103  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Wordperfect 5.1  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/261,432  
CC FILING DATE: June 17, 1994  
CC CLASSIFICATION: 800  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/003,311  
CC FILING DATE: January 12, 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Lori Y. Beardell  
CC REGISTRATION NUMBER: 34,293  
CC REFERENCE/DOCKET NUMBER: UPN-1864  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 215-568-3100  
CC TELEFAX: 215-568-3439  
CC INFORMATION FOR SEQ ID NO: 5:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 6312 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC SEQUENCE 6312 BP; 1817 A; 1106 C; 1297 G; 2092 T; 0 OTHER.  
Query Match 2.6%; Score 17; DB 1; Length 6312;  
Best Local Similarity 100.0%; Pred. No. 8.93e-01;



Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 2999 GCTGTGATGTATGT 3015  
 59 GCTGTGATGTATGT 75

RESULT 15  
 ID US-08-003-311B-3 STANDARD; DNA; UNC; 6312 BP.  
 AC xxxxxx

DE Sequence 3, Application US/08003311B  
 CC Sequence 3, Application US/08003311B  
 CC Patent No. 5444166

GENERAL INFORMATION:  
 CC APPLICANT: Ecker, Joseph R.  
 CC APPLICANT: Kiebert, Joseph J.  
 CC TITLE OF INVENTION: Constitutive Triple Response Gene  
 CC TITLE OF INVENTION: and Mutations  
 CC NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and  
 CC ADDRESSEE: No. 5444166ris  
 CC STREET: One Liberty Place - 46th Floor  
 CC City: Philadelphia  
 CC STATE: PA

COUNTRY: U.S.A.

ZIP: 19103

COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Wordperfect 5.1

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/003,311B

CC FILING DATE: January 12, 1993

CC CLASSIFICATION: 800

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 07/928,464

CC FILING DATE: August 10, 1992

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Lori Y. Beardell

CC REGISTRATION NUMBER: 34,293

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 215-568-3100

CC TELEFAX: 215-568-3439

CC INFORMATION FOR SEQ ID NO: 3:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 6312 base pairs

CC TYPE: nucleic acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: DNA (genomic)

CC SEQUENCE 6312 BP; 1817 A; 1107 C; 1297 G; 2091 T; 0 OTHER.

Query Match 2.6%; Score 17; DB 1; Length 6312;

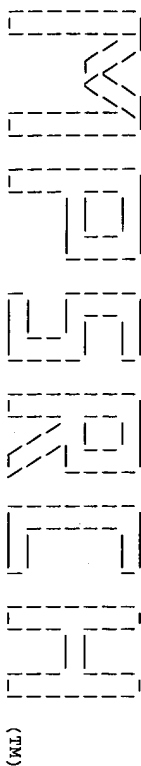
Best Local Similarity 100.0%; Pred. No. 8.93e-01;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2999 GCTGTGATGTATGT 3015  
 59 GCTGTGATGTATGT 75

Search completed: Sat Nov 27 15:06:19 1999  
 Job time : 73 secs.





Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 14:38:11 1999; MasPar time 1134.49 Seconds

Tabular output not generated. 1363.158 Million cell updates/sec

Title: >US-09-103-287-3  
Description: (1-660) from US09103287.seq  
Perfect Score: 660  
N.A. Sequence: 1 ATTTAAGATTCGATGACA.....GCGTTTAAATGTTTAA 660  
Comp: TAAATTCGATGACCTACTGT.....CGCAATATATACAAATAT

Scoring table: TABLE Jmetric

Mismatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-Processing: Minimum Match 08  
Listing first 45 summaries

Database: emb1-est58  
1:em\_est10 2:em\_est11 3:em\_est17 4:em\_est18 5:em\_est2

Database: genbank-est111  
8:gb\_est1 9:gb\_est10 10:gb\_est11 11:gb\_est12 12:gb\_est13

13:gb\_est14 14:gb\_est15 15:gb\_est16 16:gb\_est17  
17:gb\_est18 18:gb\_est19 19:gb\_est20 20:gb\_est21  
21:gb\_est22 22:gb\_est23 23:gb\_est24 24:gb\_est25  
25:gb\_est26 26:gb\_est27 27:gb\_est28 28:gb\_est29  
29:gb\_est30 30:gb\_est31 31:gb\_est32 32:gb\_est33 33:gb\_est34  
34:gb\_est35 35:gb\_est36 36:gb\_est37 37:gb\_est38 38:gb\_est39  
39:gb\_est40 40:gb\_est41 41:gb\_est42 42:gb\_est43

Statistics: Mean 8.934; Variance 1.331; scale 6.712

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	21	3.2	218	31	H49793	2.62e-08
2	21	3.2	387	19	T91899	2.62e-08
3	20	3.0	301	8	D24562	1.26e-06
4	19	2.9	196	13	AA480754	5.23e-05
5	19	2.9	264	19	C84769	5.23e-05
6	19	2.9	399	39	AA0093108	5.23e-05
7	19	2.9	497	24	A1256065	5.23e-05
8	19	2.9	614	14	C25739	5.23e-05
9	19	2.9	618	42	AQ415275	5.23e-05
10	19	2.9	674	20	C89676	5.23e-05

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
11	19	2.9	679	38	B94848	5.23e-05
12	18	2.7	72	22	AU009689	1.86e-03
13	18	2.7	82	22	AU009065	1.86e-03
14	18	2.7	143	22	AU013417	1.86e-03
15	18	2.7	143	22	AU013680	1.86e-03
16	18	2.7	154	25	A1284171	1.86e-03
17	18	2.7	154	10	AA262130	1.86e-03
18	18	2.7	300	15	C52223	1.86e-03
19	18	2.7	300	15	C51289	1.86e-03
20	18	2.7	300	15	C52048	1.86e-03
21	18	2.7	337	27	AA988199	1.86e-03
22	18	2.7	346	37	FR0014024	1.86e-03
23	18	2.7	357	15	AA569728	1.86e-03
24	18	2.7	359	38	AQ079270	1.86e-03
25	18	2.7	367	21	A1019765	1.86e-03
26	18	2.7	387	28	A1535161	1.86e-03
27	18	2.7	388	18	AA825640	1.86e-03
28	18	2.7	405	21	AA937979	1.86e-03
29	18	2.7	429	26	AA956076	1.86e-03
30	18	2.7	433	27	AA9436327	1.86e-03
31	18	2.7	435	21	AA943324	1.86e-03
32	18	2.7	436	23	A1145238	1.86e-03
33	18	2.7	460	21	A1013962	1.86e-03
34	18	2.7	483	34	AA080822	1.86e-03
35	18	2.7	484	28	A1524625	1.86e-03
36	18	2.7	489	24	A1199737	1.86e-03
37	18	2.7	493	38	B47551	1.86e-03
38	18	2.7	549	41	AO331857	1.86e-03
39	18	2.7	551	15	AA583407	1.86e-03
40	18	2.7	571	23	A1176934	1.86e-03
41	18	2.7	574	40	AQ272214	1.86e-03
42	18	2.7	633	26	A1008955	1.86e-03
43	18	2.7	638	37	AQ003798	1.86e-03
44	18	2.7	668	42	AQ423850	1.86e-03
45	18	2.7	710	42	AQ421301	1.86e-03

#### ALIGNMENTS

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	21	3.2	218	31	H49793	2.62e-08
2	21	3.2	387	19	T91899	2.62e-08
3	20	3.0	301	8	D24562	1.26e-06
4	19	2.9	196	13	AA480754	5.23e-05
5	19	2.9	264	19	C84769	5.23e-05
6	19	2.9	399	39	AA0093108	5.23e-05
7	19	2.9	497	24	A1256065	5.23e-05
8	19	2.9	614	14	C25739	5.23e-05
9	19	2.9	618	42	AQ415275	5.23e-05
10	19	2.9	674	20	C89676	5.23e-05

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
Insert Size: 1080  
High quality sequence steps: 102  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert length: 1080 Std Error: 0.00



## FEATURES

## source

Seq primer: M13RPI  
High quality sequence stop: 102.  
Location/Qualifiers

1. .218

/organism="Homo sapiens"  
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACGAGAGATTAATTAAGATCTTTTCTTTTCTTTT 3'] (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."  
/db\_xref="GDB:3798257"  
/map="11 q13.4-q23.2"  
/clone\_lib="Soares fetal liver spleen INFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
BASE COUNT 43 a 46 c 48 g 79 t 2 others  
ORIGIN

Query Match 3.2%; Score 21; DB 31; Length 218;  
Best Local Similarity 100.0%; Pred. No. 2.62e-08;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 138 TGCTGTGATGCTGATGTGGA 158  
|||||

Oy 58 TGCTGTGATGCTGATGTGGA 78  
|||||

RESULT 2 791899 387 bp mRNA EST 22-MAR-1995  
LOCUS yw2a01.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone  
DEFINITION IMAGE:116520 5', mRNA sequence.

ACCESSION T91899  
NID 9723812  
VERSION T91899.1 GI:723812  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 387)

## REFERENCE

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, B., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

The Mashu-Merck EST Project  
Unpublished (1995)

## TITLE

JOURNAL

COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1079

High quality sequence stops: 276 Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 1079 Std Error: 0.00  
Seq primer: M13RPI  
High quality sequence stop: 276.  
Location/Qualifiers

## FEATURES

## source

1. .387

/organism="Homo sapiens"  
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;

1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACGAGAGATTAATTAAGATCTTTTCTTTTCTTTT 3'] (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

/db\_xref="GDB:472137"  
/db\_xref="taxon:9606"  
/clone\_lib="Soares fetal liver spleen INFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
BASE COUNT 71 a 81 c 100 g 134 t 1 others  
ORIGIN

Query Match 3.2%; Score 21; DB 19; Length 387;  
Best Local Similarity 100.0%; Pred. No. 2.62e-08;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 139 TGCTGTGATGCTGATGTGGA 159  
|||||

Oy 58 TGCTGTGATGCTGATGTGGA 78  
|||||

RESULT 3 D24562 301 bp mRNA EST 20-JUL-1998  
LOCUS RICR2171A Rice root Oryza sativa cDNA clone R, mRNA sequence.

DEFINITION RICR2171A Rice root Oryza sativa cDNA clone R, mRNA sequence.

ACCESSION D24562  
NID 9428414  
VERSION D24562.1 GI:428414  
KEYWORDS EST.

SOURCE Oryza sativa.  
ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.

1 (bases 1 to 301)  
Minohe, Y. and Sasaki, T.

Rice cDNA from root  
Unpublished (1995)

## COMMENT

Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: tsasaki@abr.affrc.go.jp  
PROJECT "RGP"

## FEATURES

## source

Location/Qualifiers

1. .301

/organism="Oryza sativa"  
/strain="Nipponbare, sub-species Japonica"  
/note="Prepared from seedling root."  
/db\_xref="taxon:4530"  
/clone\_lib="R"  
/clone\_lib="Rice root"

BASE COUNT 87 a 68 c 67 g 79 t  
ORIGIN

Query Match 3.0%; Score 20; DB 8; Length 301;  
Best Local Similarity 100.0%; Pred. No. 1.26e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 120 AAGGACTGCTGTGATGTG 139  
|||||

Oy 51 AAGGACTGCTGTGATGTG 70  
|||||



RESULT 4  
LOCUS AA480754 196 bp mRNA EST 23-JUN-1997  
DEFINITION SMFPCA2133SK Brugia malayi microfilaria cDNA (SAM94LS-Bmf) Brugia  
malayi cDNA clone SMFPCA2133 5', mRNA sequence.  
ACCESSION AA480754  
VERSION g2209452  
KEYWORDS Brugia malayi.  
SOURCE Brugia malayi.  
ORGANISM Brugia malayi.  
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida;  
Filarioidea; Onchocercidae; Brugia.  
AUTHORS Williams, S.A.  
TITLE Genes expressed in microfilaria of Brugia malayi  
JOURNAL Unpublished (1995)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1406902.

FEATURES  
source  
1. 196  
/organism="Brugia malayi"  
/strain="TRS Labs"  
/note="Vector: Lambda Unizap XR; Site\_1: EcoR I; Site\_2:  
Xho I; lymphatic filarial nematode parasite of humans.  
mRNA was prepared from microfilariae of Brugia malayi  
isolated from birds and converted to double stranded cDNA  
using reverse transcriptase and oligo(dt) followed by  
RNase H and DNAPol I. The library had 3.5 x 10<sup>5</sup>  
independent recombinants and average insert size was 900  
base pairs. The library was constructed by Lori Saunders.  
The library is available from Dr. S.A. Williams, email  
genome@smith.edu."  
/db\_xref="taxon:6279"  
/map="19g13.3"  
/clone\_1lb="SMFPCA2133"  
/clone\_1lb="Brugia malayi microfilaria cDNA  
(SAM94LS-Bmf)"  
/lab\_host="XLI-Blue MRF"

BASE COUNT 58 a 31 c 29 g 75 t 3 others  
ORIGIN

Query Match 2.9%; Score 19; DB 13; Length 196;  
Best Local Similarity 100.0%; Pred. No. 5.23e-05;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 50 TATTCTTTCTGCTGCTGT 68  
|||||  
Cp 327 TATTCTTTCTGCTGCTGT 309

RESULT 5  
LOCUS C84769 264 bp mRNA EST 19-FEB-1998  
DEFINITION C84769 Dictyostelium discoideum SS (H. urushihara) Dictyostelium  
discoideum cDNA clone SSF674, mRNA sequence.  
ACCESSION C84769  
VERSION g2864916  
KEYWORDS Dictyostelium discoideum.  
SOURCE Dictyostelium discoideum.  
ORGANISM Eukaryota; Dictyostellida; Dictyostelium.  
REFERENCE Mizuno, H., Morio, T. and Tanaka, Y.  
AUTHORS Developmental cDNA in Dictyostelium discoideum (229)  
TITLE

JOURNAL Unpublished (1998)  
COMMENT On Apr 14, 1993 this sequence version replaced gi:693199.  
Contact: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan  
Email: d402hu@akura.cc.tsukuba.ac.jp.

FEATURES  
source  
1. 264  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone\_1lb="SSF674"  
/clone\_1lb="Dictyostelium discoideum SS (H. urushihara)"  
/dev\_stage="slug"

BASE COUNT 133 a 34 c 17 g 78 t 2 others  
ORIGIN

Query Match 2.9%; Score 19; DB 18; Length 264;  
Best Local Similarity 100.0%; Pred. No. 5.23e-05;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 TTCAATTTTATCAATTA 25  
|||||  
Cp 508 TTCAATTTTATCAATTA 490

RESULT 6  
LOCUS AQ093108 399 bp DNA GSS 26-AUG-1998  
DEFINITION HS-3020\_B1\_A12.T7 CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate-3020 Col-23 Row-B, genomic survey  
sequence.  
ACCESSION AQ093108  
VERSION g3464555  
KEYWORDS AQ093108.1 GI:3464555  
SOURCE GSS.  
ORGANISM human.  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homnidae; Homo.  
REFERENCE Mahairas, G.G., Zackrone, R.D., Smith, T., Ripton, S., Schmidt, S.,  
1 (bases 1 to 399)  
Mahairas, G.G., Zackrone, R.D., Smith, T., Ripton, S., Schmidt, S.,  
Traicoff, R., Adjian, C., Blanchard, A., West, A. and Hood, L.E.  
TITLE Construction of a Characterized Clone Resource for Genomic  
Sequencing: Generation and Preliminary Analysis of 20,000 Sequence  
Tagged Connectors  
JOURNAL Unpublished (1997)  
COMMENT

Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3020 row: B column: 23  
Class: BAC ends  
High quality sequence stop: 399.  
Location/Qualifiers  
1. 399  
/organism="Homo sapiens"  
/note="Organ: Sperm; Vector: pBelOBAC11; BAC Clones in  
E-Coli DH10B"  
/db\_xref="taxon:9606"  
/clone\_1lb="Plate-3020 Col-23 Row-B"  
/clone\_1lb="CIT Approved Human Genomic Sperm Library D"  
/sex="male"

BASE COUNT 166 a 70 c 61 g 100 t 2 others  
ORIGIN

Query Match 2.9%; Score 19; DB 39; Length 399;



Best Local Similarity 100.0%; Pred. No. 5.23e-05;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 294 CCTGCACCCATAATATAA 312  
CP 588 CCTGCACCCATAATATAA 570

RESULT 7  
LOCUS A1256065 497 bp mRNA EST 12-NOV-1998  
DEFINITION u194f07 x1 Sugano mouse liver mlia Mus musculus cDNA clone  
IMAGE:1890085.5 similar to gb:M22382 MITOCHONDRIAL MATRIX PROTEIN  
P1 PRECURSOR (HUMAN): gb:X53584 Mouse mRNA for HSP60 protein  
(MOUSE); mRNA sequence.  
ACCESSION A1256065  
NID 93863590  
VERSION A1256065.1 GI:3863590  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 497)  
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, K.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
TITLE The Mashu-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Jan 17, 1998 this sequence version replaced gi:2043817.

CONTACT: Maria M/Mouse EST Project  
Mashu-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MG1:974409  
Trace considered overall poor quality  
Seq primer: custom primer used  
High quality sequence stop: 1.  
Location/Qualifiers  
1..497  
/organism="Mus musculus"  
/strain="C57BL"  
/note="Organ: liver; Vector: pME18S-FL3; Site\_1: DraIII  
(CACTGTG); Site\_2: DraIII (CACCATGTG); 1st strand cDNA  
was primed with an oligo(dT) primer  
[ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was  
ligated to a DraIII adaptor [TGTGGCTTCTGTG], digested  
and cloned into distinct DraIII sites of the pME18S-FL3  
vector (5' site CACTGTG, 3' site CACCATGTG). XhoI should  
be used to isolate the cDNA insert. Size selection was  
performed to exclude fragments <1.5kb. Library  
constructed by Dr. Sunto Sugano (University of Tokyo  
Institute of Medical Science). Custom primers for  
sequencing: 5' end primer CTCTCTCTTAAAGCTGCG and 3' end  
primer CGACCTGACGTCGACGACA."  
/db\_xref="taxon:10090"  
/clone="IMAGE:1890085"  
/clone\_lib="Sugano mouse liver mlia"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"

BASE COUNT 140 a 120 c 98 g 136 t 3 others  
ORIGIN  
Query Match 2.9%; Score 19; DB 24; Length 497;  
Best Local Similarity 100.0%; Pred. No. 5.23e-05;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 133 TATTCAAAATTACAAAT 151  
QY 592 TATTCAAAATTACAAAT 610

RESULT 8  
LOCUS C25739 614 bp mRNA EST 29-JUL-1997  
DEFINITION C25739 Dictyostelium discoideum FC (H. Urushihara) Dictyostelium  
discoideum cDNA clone FC-AX23, mRNA sequence.  
ACCESSION C25739  
NID 92282525  
VERSION C25739.1 GI:2282525  
KEYWORDS EST.  
SOURCE Dictyostelium discoideum.  
ORGANISM Dictyostelium discoideum.  
Eukaryota; Dictyostelidia; Dictyostelium.  
REFERENCE 1 (bases 1 to 614)  
Suzuki, K., Shimizu, H. and Urushihara, H.  
TITLE Sexual cDNA in D. discoideum(970724)  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1318551.

CONTACT: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan  
Email: d402huesakura.cc.tsukuba.ac.jp.  
Location/Qualifiers  
1..614  
/organism="Dictyostelium discoideum"  
/strain="KAX3"  
/db\_xref="taxon:44689"  
/map="X27.3-28"  
/clone="FC-AX23"  
/clone\_lib="Dictyostelium discoideum FC (H. Urushihara)"  
Location/Qualifiers

BASE COUNT 231 a 84 c 94 g 205 t  
ORIGIN  
Query Match 2.9%; Score 19; DB 14; Length 614;  
Best Local Similarity 100.0%; Pred. No. 5.23e-05;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 562 TACAAGATTAAATGATTA 580  
QY 483 TACAAGATTAAATGATTA 501

RESULT 9  
LOCUS AQ415275 618 bp DNA GSS 22-MAR-1999  
DEFINITION RPCI-11-180011.TV RPCI-11 Homo sapiens genomic clone  
RPCI-11-180011, genomic survey sequence.  
ACCESSION AQ415275  
NID 94474244  
VERSION AQ415275.1 GI:4474244  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 618)  
Zhao, S., Adams, M.D., Niernan, W., Malek, J., de Jong, P. and  
Venter, J.C.  
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
Map Building  
JOURNAL Unpublished (1997)  
COMMENT Other GSSs: RPCI-11-180011.TV  
Contact: Shaying Zhao, William Niernan, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200



Fax: 301 838 0208  
 Email: hbe@ligr.org  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pliet de Jong (pliet@dejong.med.bufileo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.bufileo.edu/ordering>) or from Research Genet cs (info@resgen.com). BAC end search page: [http://www.ligr.org/ldb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.ligr.org/ldb/hungen/bac_end_search/bac_end_search.html).  
 Seq primer: 17  
 Class: BAC ends.

# FEATURES

source  
 Location/Qualifiers  
 1..618  
 /organism="Homo sapiens"  
 /note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPCI11 Human Male BAC Library"  
 /db\_xref="GDB:7569082"  
 /db\_xref="taxon:9606"  
 /clone\_1lb="RPCI-11-180011"  
 /clone\_1lb="RPCI-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 BASE COUNT 235 a 98 c 80 g 201 t 4 others  
 ORIGIN

Query Match 2.9%; Score 19; DB 42; Length 618;  
 Best Local Similarity 100.0%; Pred. No. 5.23e-05;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 253 ACAAGCATTTTAATGAA 271  
 ||||||||||||||||  
 QY 379 ACAAGCATTTTAATGAA 397

RESULT 10  
 LOCUS C89676 674 bp mRNA EST 20-APR-1998  
 DEFINITION C89676 Dictyostellium discoideum SS (H.Urushihara) Dictyostellium discoideum cDNA clone SSA606, mRNA sequence.  
 ACCESSION C89676  
 NID G3059296  
 VERSION C89676.1 GI:3059296  
 KEYWORDS EST.  
 SOURCE Dictyostellium discoideum.  
 ORGANISM Dictyostellium discoideum.  
 REFERENCE 1 (bases 1 to 674)  
 AUTHORS Yoshino, R., Morio, T. and Tanaka, Y.  
 TITLE Developmental cDNA in Dictyostellium discoideum  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2152830.

CONTACT: Hideko Urushihara  
 Institute of Biological Sciences  
 University of Tsukuba  
 3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan  
 Email: d402huesakura.cc.tsukuba.ac.jp.  
 Location/Qualifiers  
 1..674

FEATURES  
 source  
 /organism="Dictyostellium discoideum"  
 /strain="AX4"  
 /db\_xref="taxon:44689"  
 /clone\_1lb="SSA606"  
 /clone\_1lb="Dictyostellium discoideum SS (H.Urushihara)"  
 /dev\_stage="slug"  
 BASE COUNT 269 a 90 c 97 g 218 t  
 ORIGIN

Query Match 2.9%; Score 19; DB 20; Length 674;  
 Best Local Similarity 100.0%; Pred. No. 5.23e-05;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 616 TACAAGTTTAATGATA 634  
 ||||||||||||||||  
 QY 483 TACAAGTTTAATGATA 501

RESULT 11  
 LOCUS B94848 679 bp DNA GSS 24-JUN-1998  
 DEFINITION CIT-HSP-2163E2.TF CIT-HSP Homo sapiens genomic clone 2163E2, genomic survey sequence.  
 ACCESSION B94848  
 NID 92977185  
 VERSION B94848.1 GI:2977185  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 679)  
 Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and Venter, J.C.  
 Use of a random BAC End Sequence Database for Sequence-Ready Map Building  
 Unpublished (1997)  
 Other GSS: CIT-HSP-2163E2.TF  
 CONTACT: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdamas@ligr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC end search page: [http://www.ligr.org/ldb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.ligr.org/ldb/hungen/bac_end_search/bac_end_search.html)  
 Seq primer: M13-21;  
 Class: BAC ends.

FEATURES  
 source  
 Location/Qualifiers  
 1..679  
 /organism="Homo sapiens"  
 /note="Vector: pBelBAC11; Site\_1: HindIII; Site\_2: HindIII"  
 /db\_xref="taxon:9606"  
 /clone\_1lb="CIT-HSP"  
 /sex="Male"  
 /cell\_type="Sperm"  
 BASE COUNT 177 a 169 c 138 g 195 t  
 ORIGIN

Query Match 2.9%; Score 19; DB 38; Length 679;  
 Best Local Similarity 100.0%; Pred. No. 5.23e-05;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 59 TGCTTACTTAACCTTCT 77  
 ||||||||||||||||  
 Cp 421 TGCTTACTTAACCTTCT 403

RESULT 12  
 LOCUS AU009689 72 bp mRNA EST 31-JUL-1998  
 DEFINITION AU009689 Schizosaccharomyces pombe late log phase cDNA Schizosaccharomyces pombe cDNA clone spc05248, mRNA sequence.  
 ACCESSION AU009689  
 NID G3346369  
 VERSION AU009689.1 GI:3346369  
 KEYWORDS EST.  
 SOURCE fission yeast.  
 ORGANISM Schizosaccharomyces pombe  
 Eukaryota; Fungi; Ascomycota; Archiascomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomyces.  
 1 (bases 1 to 72)  
 Morimyo, M. and Mita, K.  
 Identification of expressed sequence tags of Schizosaccharomyces pombe



JOURNAL Unpublished (1998)  
COMMENT On Jan 17, 1998 this sequence version replaced gi:1900413.

Contact: Mitsuoki Morimyo  
Genome Research Group  
National Institute of Radiological Sciences  
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan  
Email: morimyo@nirs.go.jp

FEATURES  
source Location/Qualifiers

1..72  
/organism="Schizosaccharomyces pombe"  
/strain="972"

/note="Vector: M13mp19. The cDNA library of Schizosaccharomyces pombe was prepared by cloning cDNA into the SmaI site of M13mp19 DNA and the direction of DNA sequences was not always from 5' to 3'. The cDNA data of Schizosaccharomyces pombe are available for searching on the World Wide Web. (URL, <http://www.nirs.go.jp>)"

/db\_xref="taxon:4896"  
/clone\_lib="Schizosaccharomyces pombe late log phase cDNA"  
/sex="h minus"

BASE COUNT 28 a 6 c 8 g 30 t  
ORIGIN

Query Match 2.7%; Score 18; DB 22; Length 72;  
Best Local Similarity 100.0%; Pred. No. 1.86e-03;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 19 TAATTGAATATTGGAG 36  
|||||  
Cp 44 TAATTGAATATTGGAG 27

RESULT 13  
LOCUS AU009065 82 bp mRNA EST 31-JUL-1998

DEFINITION AU009065 Schizosaccharomyces pombe late log phase cDNA  
Schizosaccharomyces pombe cDNA clone spc04409, mRNA sequence.

ACCESSION AU009065  
NID 93345745

VERSION AU009065.1 GI:3345745

KEYWORDS EST.

SOURCE fission yeast.

ORGANISM Schizosaccharomyces pombe

Eukaryota; Fungi; Ascomycota; Archiascomycetes;

Schizosaccharomycetales; Schizosaccharomycetaceae;

Schizosaccharomyces.  
1 (bases 1 to 82)

1 (bases 1 to 82)  
Morimyo,M. and Mita,K.  
Identification of expressed sequence tags of Schizosaccharomyces

Pombe  
Unpublished (1998)  
On Jan 14, 1998 this sequence version replaced gi:1197110.

JOURNAL  
COMMENT

Contact: Mitsuoki Morimyo  
Genome Research Group  
National Institute of Radiological Sciences  
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan  
Email: morimyo@nirs.go.jp

FEATURES  
source Location/Qualifiers

1..82  
/organism="Schizosaccharomyces pombe"  
/strain="972"

/note="Vector: M13mp19. The cDNA library of Schizosaccharomyces pombe was prepared by cloning cDNA into the SmaI site of M13mp19 DNA and the direction of DNA sequences was not always from 5' to 3'. The cDNA data of Schizosaccharomyces pombe are available for searching on the World Wide Web. (URL, <http://www.nirs.go.jp>)"

/db\_xref="taxon:4896"  
/clone\_lib="Schizosaccharomyces pombe late log phase cDNA"  
/sex="h minus"

BASE COUNT 34 a 9 c 7 g 32 t  
ORIGIN

Query Match 2.7%; Score 18; DB 22; Length 82;  
Best Local Similarity 100.0%; Pred. No. 1.86e-03;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 37 CTCGAATATTCAAAATTA 54  
|||||  
Oy 27 CTCGAATATTCAAAATTA 44

RESULT 14  
LOCUS AU013417 143 bp mRNA EST 03-AUG-1998

DEFINITION AU013417 Schizosaccharomyces pombe late log phase cDNA  
Schizosaccharomyces pombe cDNA clone spc08123, mRNA sequence.

ACCESSION AU013417  
NID 93368208

VERSION AU013417.1 GI:3368208

KEYWORDS EST.

SOURCE fission yeast.

ORGANISM Schizosaccharomyces pombe

Eukaryota; Fungi; Ascomycota; Archiascomycetes;

Schizosaccharomycetales; Schizosaccharomycetaceae;

Schizosaccharomyces.  
1 (bases 1 to 143)  
Morimyo,M. and Mita,K.  
Identification of expressed sequence tags of Schizosaccharomyces

Pombe  
Unpublished (1998)  
On Jan 19, 1998 this sequence version replaced gi:2285245.

JOURNAL  
COMMENT

Contact: Mitsuoki Morimyo  
Genome Research Group  
National Institute of Radiological Sciences  
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan  
Email: morimyo@nirs.go.jp

FEATURES  
source Location/Qualifiers

1..143  
/organism="Schizosaccharomyces pombe"  
/strain="972"

/note="Vector: M13mp19. The cDNA library of Schizosaccharomyces pombe was prepared by cloning cDNA into the SmaI site of M13mp19 DNA and the direction of DNA sequences was not always from 5' to 3'. The cDNA data of Schizosaccharomyces pombe are available for searching on the World Wide Web. (URL, <http://www.nirs.go.jp>)"

/db\_xref="taxon:4896"  
/clone\_lib="Schizosaccharomyces pombe late log phase cDNA"  
/sex="h minus"

BASE COUNT 53 a 20 c 16 g 54 t  
ORIGIN

Query Match 2.7%; Score 18; DB 22; Length 143;  
Best Local Similarity 100.0%; Pred. No. 1.86e-03;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 92 CTCGAATATTCAAAATTA 109  
|||||  
Oy 27 CTCGAATATTCAAAATTA 44

RESULT 15  
LOCUS AU013680 143 bp mRNA EST 03-AUG-1998

DEFINITION AU013680 Schizosaccharomyces pombe late log phase cDNA  
Schizosaccharomyces pombe cDNA clone spc08497, mRNA sequence.

ACCESSION AU013680  
NID 93368471

VERSION AU013680.1 GI:3368471

KEYWORDS EST.

SOURCE fission yeast.

ORGANISM Schizosaccharomyces pombe



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Fungi; Ascomycota; Archiascomycetes;  
Schizosaccharomycetales; Schizosaccharomycetaceae;  
Schizosaccharomyces.  
1 (bases 1 to 143)  
Morimyo, M. and Mita, K.  
Identification of expressed sequence tags of Schizosaccharomyces  
pombe  
Unpublished (1998)  
On Jan 19, 1998 this sequence version replaced gi:2286292.

CONTACT: Mitsuki Morimyo  
Genome Research Group  
National Institute of Radiological Sciences  
9-1, Anagawa-4-chome, Inage-Ku, Chiba 263-8555, Japan  
Email: morimyo@nirs.go.jp.  
Location/Qualifiers

FEATURES  
source  
1..143  
/organism="Schizosaccharomyces pombe"  
/strain="972"  
/note="Vector: M13mp19; The cDNA library of  
Schizosaccharomyces pombe was prepared by cloning cDNA  
into the SmaI site of M13mp19 DNA and the direction of DNA  
sequences was not always from 5' to 3'. The cDNA data of  
Schizosaccharomyces pombe are available for searching on  
the World Wide Web. (URL, <http://www.nirs.go.jp>)"  
/db\_xref="taxon:4896"  
/clone\_lib="SPC08497"  
/clone="Schizosaccharomyces pombe late log phase cDNA"  
/sex="h minus"  
BASE COUNT 53 a 20 c 16 g 54 t  
ORIGIN

Query Match 2.7%; Score 18; DB 22; Length 143;  
Best Local Similarity 100.0%; Pred. No. 1.86e-03;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 92 CTCGAATATTTCGAATT 109  
|||  
0y 27 CTCGAATATTTCGAATT 44

Search completed: Sat Nov 27 15:00:22 1999  
Job time : 1331 secs.



**THIS PAGE BLANK (USPTO)**



\*\*\*\*\*  
 WISE (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_un n.a. - n.a. database search, using Smith-Waterman algorithm  
 Run on: Sat Nov 27 15:06:36 1999; MasPar time 76.40 Seconds  
 Tabular output not generated. 689.358 Million cell updates/sec

Title: >US-09-103-287-5  
 Description: (1-19) from US09103287.seq  
 Perfect Score: 19  
 N.A. Sequence: 1 CTTGATTATGACAGC 19  
 Comp: GAAGTAATTACTGCTAGC

Scoring table: TABLE jmetric  
 Gap 60  
 Mismatch SPD: Dbase 0; Query 0

Searched: 646147 segs, 138595363 bases x 2

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: emb158  
 1:em\_ba1 2:em\_ba2 3:em\_fun 4:em\_htg 5:em\_hum1 6:em\_hum2  
 7:em\_in 8:em\_com 9:em\_or 10:em\_ov 11:em\_pat 12:em\_ph  
 13:em\_pl 14:em\_ro 15:em\_sts 16:em\_vl  
 Database: genbank111  
 17:gb\_ba1 18:gb\_ba2 19:gb\_htg1 20:gb\_htg2 21:gb\_in1  
 22:gb\_in2 23:gb\_com 24:gb\_ov 25:gb\_pat 26:gb\_ph 27:gb\_pl1  
 28:gb\_pl2 29:gb\_pl1 30:gb\_pl2 31:gb\_pl3 32:gb\_ro  
 33:gb\_st 34:gb\_sts 35:gb\_sy 36:gb\_un 37:gb\_vl

Statistics: Mean 5.848; Variance 2.402; scale 2.435

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	16	84.2	3762	17	BOU15178	Bacteroides ovatus ara 2.28e+00
2	15	78.9	39489	30	HSU69570	Human XP22 cosmid U27H 1.14e+01
3	15	78.9	41351	21	CELF42C5	Caenorhabditis elegans 1.14e+01
4	15	78.9	199917	19	AC006755	Caenorhabditis elegans 1.14e+01
5	14	73.7	260	17	PHU59225	Pasteurella haemolytic 5.42e+01
6	14	73.7	260	17	PHU59226	Pasteurella haemolytic 5.42e+01
7	14	73.7	436	34	G47706	225840_1 zebrafish AB 5.42e+01
8	14	73.7	1515	22	AF073334	Schistosoma japonicum 5.42e+01
9	14	73.7	1608	21	DDU96316	Schistosoma japonicum 5.42e+01
10	14	73.7	1724	27	D89250	Schistosoma japonicum 5.42e+01
11	14	73.7	1964	32	MUSTPX101	Schistosoma japonicum 5.42e+01
12	14	73.7	2316	28	AT048698	Mouse testis-specific 5.42e+01
13	14	73.7	2609	28	ANNA22	Arabidopsis thaliana P 5.42e+01
14	14	73.7	2609	28	ANNA22	A.nidus mitochondrial 5.42e+01

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
14	14	73.7	2892	17	BACHSMR	Bacillus subtilis 15B8 5.42e+01
15	14	73.7	27227	27	SPAC22H10	S.pombe chromosome I c 5.42e+01
16	14	73.7	28735	21	CELC02D5	C. elegans cosmid C02D 5.42e+01
17	14	73.7	36414	21	CELM01C8	Caenorhabditis elegans 5.42e+01
18	14	73.7	37132	22	SPCC1919	S.pombe chromosome III 5.42e+01
19	14	73.7	38155	22	CELF11C7	Caenorhabditis elegans 5.42e+01
20	14	73.7	41883	21	CELD2096	Caenorhabditis elegans 5.42e+01
21	14	73.7	42730	21	CEC23H4	Caenorhabditis elegans 5.42e+01
22	14	73.7	42824	31	AC004207	Homo sapiens clone UMG 5.42e+01
23	14	73.7	83343	20	AC005472	Drosophila melanogaster 5.42e+01
24	14	73.7	87841	27	AB005247	Arabidopsis thaliana g 5.42e+01
25	14	73.7	95274	29	HS65019	Human DNA sequence fto 5.42e+01
26	14	73.7	135356	31	AC004926	Human DNA sequence fto 5.42e+01
27	14	73.7	166180	19	HS695020	Human DNA sequence *** 5.42e+01
28	14	73.7	179485	31	AC005939	Human DNA sequence *** 5.42e+01
29	14	73.7	180551	30	HUAC004158	Homo sapiens chromosome 5.42e+01
30	14	73.7	186510	29	HS451B15	Human DNA sequence fto 5.42e+01
31	13	68.4	743	25	188004	Sequence 4 from Patent 2.42e+02
32	13	68.4	1001	37	FBNYVC5	Faba bean necrotic yel 2.42e+02
33	13	68.4	1707	25	103379	Sequence 1 from Patent 2.42e+02
34	13	68.4	2017	27	SCYOL031C	S.cerevisiae chromosome 2.42e+02
35	13	68.4	3705	27	SCYBR114W	S.cerevisiae chromosome 2.42e+02
36	13	68.4	5430	24	AB007446	Gallus gallus mraA for 2.42e+02
37	13	68.4	8949	27	SCMDLIX	S.cerevisiae YSA1, SSN 2.42e+02
38	13	68.4	37704	21	CEFC369	Caenorhabditis elegans 2.42e+02
39	13	68.4	39871	27	CAC38F10	C.albicans cosmid Ca38 2.42e+02
40	13	68.4	53264	19	HS315G5	Human DNA sequence *** 2.42e+02
41	13	68.4	60050	22	AC002441	Drosophila melanogaster 2.42e+02
42	13	68.4	86621	30	AC003675	Human Chromosome 11p15 2.42e+02
43	13	68.4	122577	19	AC004825	Homo sapiens clone DJ0 2.42e+02
44	13	68.4	138343	19	AC004907	Homo sapiens clone DJ0 2.42e+02
45	13	68.4	200965	20	AC006740	Caenorhabditis elegans 2.42e+02

## ALIGNMENTS

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	16	84.2	3762	17	BOU15178	Bacteroides ovatus ara 2.28e+00
2	15	78.9	39489	30	HSU69570	Human XP22 cosmid U27H 1.14e+01
3	15	78.9	41351	21	CELF42C5	Caenorhabditis elegans 1.14e+01
4	15	78.9	199917	19	AC006755	Caenorhabditis elegans 1.14e+01
5	14	73.7	260	17	PHU59225	Pasteurella haemolytic 5.42e+01
6	14	73.7	260	17	PHU59226	Pasteurella haemolytic 5.42e+01
7	14	73.7	436	34	G47706	225840_1 zebrafish AB 5.42e+01
8	14	73.7	1515	22	AF073334	Schistosoma japonicum 5.42e+01
9	14	73.7	1608	21	DDU96316	Schistosoma japonicum 5.42e+01
10	14	73.7	1724	27	D89250	Schistosoma japonicum 5.42e+01
11	14	73.7	1964	32	MUSTPX101	Schistosoma japonicum 5.42e+01
12	14	73.7	2316	28	AT048698	Mouse testis-specific 5.42e+01
13	14	73.7	2609	28	ANNA22	Arabidopsis thaliana P 5.42e+01
14	14	73.7	2609	28	ANNA22	A.nidus mitochondrial 5.42e+01

## FEATURES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	16	84.2	3762	17	BOU15178	Bacteroides ovatus ara 2.28e+00
2	15	78.9	39489	30	HSU69570	Human XP22 cosmid U27H 1.14e+01
3	15	78.9	41351	21	CELF42C5	Caenorhabditis elegans 1.14e+01
4	15	78.9	199917	19	AC006755	Caenorhabditis elegans 1.14e+01
5	14	73.7	260	17	PHU59225	Pasteurella haemolytic 5.42e+01
6	14	73.7	260	17	PHU59226	Pasteurella haemolytic 5.42e+01
7	14	73.7	436	34	G47706	225840_1 zebrafish AB 5.42e+01
8	14	73.7	1515	22	AF073334	Schistosoma japonicum 5.42e+01
9	14	73.7	1608	21	DDU96316	Schistosoma japonicum 5.42e+01
10	14	73.7	1724	27	D89250	Schistosoma japonicum 5.42e+01
11	14	73.7	1964	32	MUSTPX101	Schistosoma japonicum 5.42e+01
12	14	73.7	2316	28	AT048698	Mouse testis-specific 5.42e+01
13	14	73.7	2609	28	ANNA22	Arabidopsis thaliana P 5.42e+01
14	14	73.7	2609	28	ANNA22	A.nidus mitochondrial 5.42e+01



/translation-"MRYANLVLALSTLALHAQNEIVTOKKIGAEIOTPTGL  
FFEDINYAADGGLYAEIVKNSFEFQHLGKMTYGVSLMNDGPERNPHYVRLSDP  
GHAHKTGLDNEGFEGIVKKGSEYRSVWARLPQSGTETLELIVDTOSMERAL  
VAGNLITDSKDMKKYOMILKPGSTHPSVRLFTSGTVDLHSLFEPDTMKGHEN  
GLKRDLAOLADLHPGVEFRPGGCIYEGTDLERYDMKRSVGVENPLNENMOYTF  
THREPPDYOSYGLGFEYFLLSEKNAALPLINGLSIQYONNPKAIVACDDN  
YIDADLULIFANGNVNTGKATRADGHPAPNKLKIGIGNECWKEPERLEPFLK  
AIRRAHEIKIVGSSGPNSEKDFDYLMPMKRLKVDLVEHEFYRPSWFLAOGARD  
NYDRKGFKEVAGEACGKCKGNHHALEAFMTGLERNADIVMATAYALFAHY  
EGWQMRDMMIFDNLNSVRTSYVQOLYKNGTNYVLPITMKNKIVYTGAEONGLEA  
SAYDKGKMLIVKANTSATIOPISLNEFKKODVLSGRCIKLRSLDLDNLTLE  
QPGYIVQEPVSEGVFTTELEPTFAVYKFTTK"

BASE COUNT 1098 a 834 c 782 g 1045 t 3 others

Query Match 84.2%; Score 16; DB 17; Length 3762;  
Best Local Similarity 100.0%; Pred. NO. 2.28e+00;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 356 TCATTAAATGAACGATG 371  
|||||  
3 TCATTAAATGAACGATG 18

QY 3 TCATTAAATGAACGATG 18

RESULT 2 HSU69570 39489 bp DNA PRI 07-FEB-1997  
LOCUS Human Xp22 cosmid U27H1, complete sequence.  
DEFINITION U69570  
ACCESSION g1825473  
NID 069570.1 GI:1825473  
VERSION HTG.  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 39489)  
AUTHORS Lu, J. and Gibbs, R. A.  
TITLE Direct Submission  
JOURNAL Submitted (04-SEP-1996) Molecular and Human Genetics, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
COMMENT On Feb 7, 1997 this sequence version replaced gi:1552535.  
FEATURES  
Source  
1. .39489 Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="Xp22"  
/clone="10E6"  
/clone\_1b="LNL X chromosome library L10XNC01"  
350. .798  
repeat\_region /rpt\_family="L1MA2"  
420. .608  
repeat\_region /rpt\_family="L1MA10"  
complement(955. .1110)  
repeat\_region /rpt\_family="Alu"  
1834. .2117  
repeat\_region /rpt\_family="Alu"  
2147. .2419  
repeat\_region /rpt\_family="Alu"  
4404. .4573  
STS /db\_xref="dbSTS:G02051"  
complement(4928. .5080)  
repeat\_region /rpt\_family="L1"  
5642. .5731  
misc-feature /note="similar to ESTs with GenBank Accession Numbers  
W27937 and N29255"  
/note="Region: 39910"  
/db\_xref="dbEST:W27937"  
/db\_xref="dbEST:N29255"  
6140. .6427  
repeat\_region /rpt\_family="Alu"  
7504. .7578  
repeat\_region /rpt\_family="MIR"

repeat\_region complement(9106. .9726)  
/rpt\_family="MERIA"  
repeat\_region complement(10925. .10987)  
/rpt\_family="MLT1f"  
misc-feature /note="similar to ESTs with GenBank Accession Numbers  
W27937, N29126, N20542."  
12926. .12996  
repeat\_region /note="Region: 39910"  
complement(14647. .14754)  
repeat\_region /rpt\_family="MIR"  
15217. .15311  
repeat\_region /rpt\_family="MERS"  
15783. .16060  
repeat\_region /rpt\_family="Alu"  
misc-feature /note="similar to EST with GenBank Accession Number  
N47571"  
/note="Region: YY91f09.r1"  
17958. .18376  
misc-feature /note="similar to ESTs with GenBank Accession Numbers  
W29126, N47571, W27937 and N20542"  
/note="Region: YY91f09.s1"  
complement(20857. .21139)  
repeat\_region /rpt\_family="Alu"  
complement(28438. .29040)  
repeat\_region /rpt\_family="L1ME2"  
32018. .32074  
repeat\_region /rpt\_family="MERS"  
complement(33579. .33732)  
repeat\_region /rpt\_family="L1MD1"  
complement(35236. .35487)  
repeat\_region /rpt\_family="Alu"  
36873. .37150  
repeat\_region /rpt\_family="Alu"  
37761. .38007  
repeat\_region /rpt\_family="Alu"  
repeat\_region complement(39161. .39489)  
/rpt\_family="MLT1b"  
BASE COUNT 12027 a 7669 c 7510 g 12283 t

Query Match 78.9%; Score 15; DB 30; Length 39489;  
Best Local Similarity 100.0%; Pred. NO. 1.14e+01;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22194 GCATCGTTCATTAT 22208  
|||||  
19 GCATCGTTCATTAT 5

CP 19 GCATCGTTCATTAT 5

RESULT 3 CELF42C5 41351 bp DNA INV 17-NOV-1995  
LOCUS Caenorhabditis elegans cosmid F42C5.  
DEFINITION U40799.  
ACCESSION g1065935  
NID 040799.1 GI:1065935  
VERSION HTG.  
KEYWORDS Caenorhabditis elegans strain-Bristol N2.  
SOURCE Caenorhabditis elegans  
ORGANISM Eukaryote; mitochondrial eukaryotes; Metazoa; Nematoda;  
Secernentea; Rhabdita; Rhabditida; Rhabditina; Rhabditioidea;  
Rhabditidae; Caenorhabditis.  
REFERENCE 1 (bases 1 to 41351)  
AUTHORS Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berke, M.,  
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,  
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,  
Fulton, L., Gardner, A., Green, P., Hawkins, J., Hillier, L., Jler, M.,  
Johnston, L., Jones, M., Kersey, J., Kirsten, J., Laister, N.,  
Latreille, P., Lightning, J., Lloyd, C., McMuray, A., Mortimore, B.,  
O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,  
Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E.,  
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,  
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,



TITLE Wilkinson-Sprat, J. and Wohldman, P.  
2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*  
JOURNAL Nature 368 (6466), 32-38 (1994)  
MEDLINE 94150718  
REFERENCE 2 (bases 1 to 41351)  
AUTHORS Du, Z.  
TITLE The sequence of *C. elegans* cosmid F42C5  
JOURNAL Unpublished (1995)  
REFERENCE 3 (bases 1 to 41351)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-1995) Robert Waterston  
COMMENT Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
e-mail: tw@emulcode.wustl.edu and jesse@sanger.ac.uk  
NEIGHBORING COSMID INFORMATION:  
The 5' cosmid is C06E4, 200 bp overlap; 3' end lies in a gap followed by the cosmid T1283. Actual start of this cosmid is at base position 1 of CELF42C5; actual end is at base position 41351 of CELF42C5.

## NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFINDER (P. Green and L. Hillier, ms in preparation).

## FEATURES

## source

Location/Qualifiers

1..41351

/organism="Caenorhabditis elegans"

/db\_xref="taxon:6239"

/map="IV"

/complement(979..2164)

/gene="F42C5.3"

/complement(join(979..1116,1265..1333,1382..1501,1545..1778,1973..2164))

/gene="F42C5.3"

/codon\_start=1

/evidence=not\_experimental

/protein\_id="AA81479.1"

/db\_xref="PID:91065936"

/db\_xref="GI:1065936"

/translation="MEDHSEKMDRLRQAANVSEKETFDKRRRETLLEKHKDEL  
EKKNKEDDLREONNNLNKNDLDOOLAGNELNNKYTEISNQNSRLQIKQOEEDIL  
VKIKLIDSKYVTDYKVVYQVLYRNTLDENSNKSDVLEKIDPLIYKKELEKIL  
TAKKILGKQKRTTSDSKVINDSLTELMREFEDDILIEKTTIKKNSAEOSEFLTK  
MDETIDKYNEMVNEPLGLQKSEVPIHQAIO"

/complement(2977..6924)

/gene="F42C5.4"

/complement(join(2977..3636,3722..3835,3885..3997,4245..4323,4369..5030,5213..5405,5480..5670,5778..5846,6667..6924))

/gene="F42C5.4"

/note="short region of similarity to *S. cerevisiae* nuclear polyadenylated RNA-binding protein NAB3 (SP:P38996)"

/codon\_start=1

/evidence=not\_experimental

/protein\_id="AA81480.1"

/db\_xref="PID:91065937"

/db\_xref="GI:1065937"

/translation="MEMKREARLESHGSIATEMLHEVSLVSGSGDGLDRREFDN  
DKRPVCSNHSICSCDFTECEPCKEPIVRSNVOARKIVHGFKNPMILIEK  
NMTANGSETCCHNDAKKPRFCVTCGQKQYILMKCRPKRGIERDDEBEKRR  
QSETRSKKTLIVNEFLPKSNWTDKMARVFGCPVNMNQLLCAQCCFHHEDH  
ALRTSEQLMYIDADWNRSAKLIVSDFIWELSRSKCLIKTMRLHRTCEKLIHFVK  
FYTYEPDSRSHVGRMRSDYKFEESLSRKNVDSVREYDELIOISLQIKLYLY  
HDDKCDIDIMNMHKSIFGNQYKLEFVYINLEDEBEIKDCPLTMDKONIONKEEV  
SKHJLRIYAFCAQNTTEKGVHFLPDNNHRCVCPNCSITLGIKSNMLYKSGFC  
GSAFPEVGLYREFNEDYNILEYGLYKRNQVDFLEMMWRCEPSEHLCSCSSYSE  
ELICVCFELDRHKALTKPKTIENSIGKRNLSLEFNISGNDPFIHQSDCKRRL  
HSNWHASQDSSIPERHMKGTWRGCHNIIIRILRSKCKCYANADECEVNSITLKSIR

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

NYERAKATWGLIFRILKSGIENITCKIRGENMLNTYGLKSKCRFFINSKEEI  
NEENRNSNEINKELDKIQKNGITTI  
/complement(7399..8928)  
/gene="F42C5.5"  
/complement(join(7399..7449,7500..7864,7915..8305,8764..8928))  
/gene="F42C5.5"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AA81481.1"  
/db\_xref="PID:91065938"  
/db\_xref="GI:1065938"

/translation="WKTYREESQOTTVGEVDFVHLQTRGETTYQODIDNMDL  
EKIERFIRNGENKRENTVTCEDILDRNDFEYHANFRTPYGNFLISOAPDNSTNA  
DHLHLMYINQTVCLVPEAGGYEVPKGEFTSFQKQIKTIGFDEKQGVY  
OCEKSYAPKRLNRMVYIICDPSVGSIRSPQOYIMETWMAFEETNAIEEVAL  
ADATTYVHICAGRCASFVAVAQKOLLETFQSAMETWEMRRRAHCTGRH  
DFESSVTFEAFVCSAVSETENYAKTMEIVADLVKKKEKND"

19671..23147

/gene="F42C5.2"

/complement(join(19671..19844,21175..21253,21319..21427,21912..22163,22382..22873,22922..22974,23026..23147))

/gene="F42C5.2"

/note="similar to *Mu*-typ opiod and somatostatin receptors"

/codon\_start=1

/evidence=not\_experimental

/protein\_id="AA81482.1"

/db\_xref="PID:91065939"

/db\_xref="GI:1065939"

/translation="WTKRWISIKYANGFHOICKEDFEDVYANNKCVTLVDRLKL  
ELEKQSGSFNGSILEIFSLYLFPLDKSFPEMFIEENDIDSFAKGFMTVYIV  
CLVGPNGMNIYKILFRALWGSAGVOLTVQSRSTYFALACSMFLTPATASY  
NYHGWINGSAACYIIRSIETIFAKFLSVLTVMSLEYIIVCTRLRIYANMSLVP  
LAVGTIFGVLPVTIIHFYLOHFSVPEDVTVTCPLMSNEVNLFPQYTFVGGFI  
EPATMFACVIMLVNHRVTKYKMRRAQTTLAKGKERHYSEVKKSIMRAVHPFC  
WAPWGFMLPNYITQIDQFLGDNDEESGESIFLYCRLVSNCLPTINAGMWLV  
ALLNVYVRKHLYINOPKRRKFTLKNFNTSSNC"

/complement(23283..24752)

/gene="F42C5.6"

/complement(join(23283..23477,23525..23665,23714..23748,24683..24752))

/gene="F42C5.6"

/codon\_start=1

/evidence=not\_experimental

/protein\_id="AA81483.1"

/db\_xref="PID:91065940"

/db\_xref="GI:1065940"

/translation="WHSGLCCITYDHSLSLSDSSPEOTEGSIGTTVAKIESDA  
IWCPTHRVGNCPETTLISFYKCGHLKKECCFLRWIVIMIIIVLPICLIVLPASTFL  
IRRICSSSQPTGAPSAADRVRLOLAEDSTFRDSDVMVL"

/complement(24822..26102)

/gene="F42C5.7"

/complement(join(24822..25001,25049..25120,25175..25555,25628..25703,26059..26102))

/gene="F42C5.7"

/codon\_start=1

/evidence=not\_experimental

/protein\_id="AA81484.1"

/db\_xref="PID:91065941"

/db\_xref="GI:1065941"

/translation="MEAVPKDVRYMAKRLRRREITASETGYVNFCAKLRVMLLP  
LSILFSLIASSTFGGCSMGPPCPAPPPAPLPAPPPPPICPPCPPPPPMC  
PP  
HRTKSHKFTTEDPEEDDCTSKLRRMENMNDPSISRAIOKAYEEMFGKFN  
ICARGDESVATYETICVANDVYCIAFRPM"

/complement(27841..28656)

/gene="F42C5.8"

/complement(join(27841..27950,27997..28302,28353..28452,28546..28656))

/gene="F42C5.8"

/note="coded for by *C. elegans* cDNA yk516.5; coded for by *C. elegans* cDNA yk107e3.5; coded for by *C. elegans* cDNA yk103h6.5; coded for by *C. elegans* cDNA yk55e5.5; coded



for by C. elegans CDNA yk78a4.5; coded for by C. elegans CDNA yk107e3.3; coded for by C. elegans CDNA yk103b6.3; coded for by C. elegans CDNA yk55e5.3; coded for by C. elegans CDNA yk78a4.3; similar to 40S ribosomal protein 58."

protein\_id="AAA81485.1"  
/db\_xref="PID:91065942"  
/db\_xref="GI:1065943"  
/translation="MGISRDSWHRKRYTGATOPVPHKKRKEFGPRANFKIGAHVR  
LVRTGNEKRYALRLDGSNFSMASEOTTRTIVDTYVNTATNNEVTKTLVGAII  
SVDAPEFROWEAHALPLARKNAKLSEDNALIKKRSHTMKYEROKTAVDA  
LLIBQFNTGRLARISSSPGVGVANGIILEKEIDFLRKIRAKKAK"  
28209..29786  
/gene="F42C5.1"  
join(28209..28440,29481..29696,29749..29786)  
/gene="F42C5.1"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAA81486.1"  
/db\_xref="PID:91065943"  
/db\_xref="GI:1065943"  
/translation="MMAPLTRYLVRTNSLFLVAVYVSTIRVLRVSEKSEYERFIKYN  
MEYTAQPRRSRSGELCTPRFLHCEPDRGSRPKMTGAVQCTLEAVEECRS  
LQWRLMRSELKFRCSISNFFFSWAVYLSKICVIFPMTDVNLTIPIGRDAPPA  
K"  
complement(29490..31891)  
/gene="F42C5.9"  
complement(join(29490..29582,29645..29792,29842..30089,  
30136..30396,30445..30934,31019..31244,31697..31891))  
/gene="F42C5.9"  
/note=coded for by C. elegans CDNA CESA62F; coded for by

...  
Note: remainder of annotations omitted.

Query Match 78.9%; Score 15; DB 21; Length 41351;  
Best Local Similarity 100.0%; Pred. No. 1.14e+01;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11954 TCGTTCATTATGAA 11968  
|||||  
Cp 16 TCGTTCATTATGAA 2

RESULT 4  
LOCUS AC006755 199917 bp DNA HTG 23-FEB-1999  
DEFINITION Caenorhabditis elegans clone Y40C5, WORKING DRAFT SEQUENCE, 1  
ACCESSION AC006755  
NID 94263183  
VERSION AC006755.1 GI:4263183  
KEYWORDS HTG; HTGS; PHASE1.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;  
Rhabditina; Rhabditidae; Rhabditidae; Felodierinae; Caenorhabditis.  
1 (bases 1 to 199917)  
Waterston, R.H.  
The sequence of Caenorhabditis elegans clone  
Unpublished  
2 (bases 1 to 199917)  
Waterston, R.H.  
Direct Submission  
Submitted (23-FEB-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

COMMENT  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will

\* be preserved.  
1 199917: contig of 199917 bp in length.  
Location/Qualifiers  
1..199917  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/clone="Y40C5"

BASE COUNT 63887 a 33980 c 34366 g 65836 t 1848 others  
ORIGIN

Query Match 78.9%; Score 15; DB 19; Length 199917;  
Best Local Similarity 100.0%; Pred. No. 1.14e+01;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 25332 TCGTTCATTATGAA 25346  
|||||  
Cp 16 TCGTTCATTATGAA 2

RESULT 5  
LOCUS PHU59225 260 bp DNA BCT 17-DEC-1996  
DEFINITION Pasteurella haemolytica serotype A2 [Cu,Zn]-superoxide dismutase  
(sodC) gene, partial cds.  
ACCESSION U59225  
NID 91389830  
VERSION U59225.1 GI:1389830  
KEYWORDS  
SOURCE Pasteurella haemolytica.  
ORGANISM Pasteurella haemolytica  
Eubacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Pasteurella.  
1 (bases 1 to 260)  
Lainson, F.A., Thomson, N., Rowe, H.A., Langford, P.R., Aitchison, K.D.,  
Donachie, W. and Kroll, J.S.  
Occurrence of [copper, zinc]-cofactored superoxide dismutase in  
Pasteurella haemolytica and its serotype distribution  
FEMS Microbiol. Lett. 142 (1), 11-17 (1996)  
96327137  
REFERENCE 2 (bases 1 to 260)  
Lainson, F.A.  
Direct Submission  
Submitted (24-MAY-1996) Moredun Research Institute, Microbiology,  
408 Gilmerston Road, Edinburgh, Scotland, UK, EH17 7UH  
JOURNAL  
TITLE  
AUTHORS  
MEDLINE  
JOURNAL  
FEATURES  
SOURCE  
1..260  
Location/Qualifiers  
/organism="Pasteurella haemolytica"  
/serotype="A2"  
/specific\_host="sheep"  
/db\_xref="taxon:746"  
/clone="pMT1"  
1..260  
/gene="sodC"  
<1..>260  
/gene="sodC"  
/function="detoxification of oxygen radicals"  
/translation="detoxification of oxygen radicals"  
/codon\_start=3  
/transl\_table=1  
/evidence=experimental  
/product="[Cu,Zn]-superoxide dismutase"  
/protein\_id="AA83871.1"  
/db\_xref="PID:91389831"  
/db\_xref="GI:1389831"  
/translation="AKKGGKLTAGLAAGGHWNPKNKAPHHGFPSDPAHLGLDLPALTV  
LHDSTNPVLAPELRKRLDEIKGRSLMHEGCDNHSHPAPL"

BASE COUNT 79 a 69 c 54 g 58 t  
ORIGIN

Query Match 73.7%; Score 14; DB 17; Length 260;  
Best Local Similarity 100.0%; Pred. No. 5.42e+01;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 203 TCGTTCATTATGAA 216  
|||||



CP 16 TCGTTCATTATGA 3

RESULT 6 PHU59226 260 bp DNA BCT 17-DEC-1996  
 LOCUS Pasteurella haemolytica serotype A7 [cu,zn]-superoxide dismutase  
 DEFINITION (sodC) gene, partial cds.  
 ACCESSION U59226  
 MID 91389832  
 VERSION U59226.1 GI:1389832  
 KEYWORDS  
 SOURCE Pasteurella haemolytica.  
 ORGANISM Pasteurella haemolytica.  
 Eubacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 Pasteurella.

REFERENCE 1 (bases 1 to 260)  
 AUTHORS Lainson, F.A., Thomson, N., Rowe, H.A., Langford, P.R., Atchison, K.D.,  
 TITLE Donachie, W. and Kroll, J.S. Occurrence of [copper, zinc]-cofactored superoxide dismutase in  
 JOURNAL Pasteurella haemolytica and its serotype distribution  
 MEDLINE FEMS Microbiol. Lett. 142 (1), 11-17 (1996).  
 REFERENCE 96327137  
 AUTHORS Lainson, F.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-MAY-1996) Moredun Research Institute, Microbiology,  
 408 Gilmerton Road, Edinburgh, Scotland, UK, EH17 7JH  
 LOCATION/Qualifiers

FEATURES  
 source 1..260  
 /organism="Pasteurella haemolytica"  
 /serotype="A7"  
 /specific\_host="sheep"  
 /db\_xref="taxon:746"  
 /clone="PNT2"  
 1..260  
 /gene="sodC"  
 <1..>260  
 /gene="sodC"  
 /function="detoxification of oxygen radicals"  
 /codon\_start=3  
 /transl\_table=11  
 /evidence=experimental  
 /product="[Cu,Zn]-superoxide dismutase"  
 /protein\_id="AB38772.1"  
 /db\_xref="PIR:Q1389833"  
 /db\_xref="GI:1389833"  
 /translation="AKKGGKLTAGIAGCHNPNKAPRHGCPMSDAILGLPALTY  
 LHDGISTNVLAFLRKLEIKRSLIMHEGGDNHSDHPAPL"

BASE COUNT 79 a 67 c 54 g 60 t

ORIGIN

Query Match 73.7% Score 14; DB 17; Length 260;  
 Best Local Similarity 100.0%; Pred. No. 5.42e+01;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 203 TCGTTCATTATGA 216  
 |||||  
 Cp 16 TCGTTCATTATGA 3

RESULT 7 647706 436 bp DNA STS 23-MAR-1999  
 LOCUS 225840.1 zebrafish AB Danio rerio STS genomic clone 225840 5',  
 DEFINITION sequence tagged site.  
 ACCESSION G47706  
 MID 94494314  
 VERSION G47706.1 GI:4494314  
 KEYWORDS STS.  
 SOURCE zebrafish.  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
 Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
 Cyprinidae; Cyprinidae; Rasbora; Danio.

REFERENCE 1 (bases 1 to 436)  
 AUTHORS Zinili, J., Jackson, D.G. and Fishman, M.C.  
 TITLE A zebrafish microsatellite map  
 JOURNAL Unpublished (1999)  
 COMMENT

Contact: Mark C. Fishman  
 Cardiovascular Research Center  
 Massachusetts General Hospital  
 Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA  
 Fax: 6177265806  
 Email: fishman@mcg.harvard.edu  
 http://zebrafish.mcg.harvard.edu  
 Primer A: TTGGAATGACGTGCA  
 Primer B: TGTGCTGTGTCATGTGCA  
 STS size: 245  
 PCR Profile:  
 Presoak: 94 degrees C for 5.0 minutes  
 Denaturation: 94 degrees C for 1.0 minute  
 Annealing: 58 degrees C for 1.0 minute  
 Polymerization: 72 degrees C for 1.5 minute  
 PCR Cycles: 27  
 Thermal Cycler: MJ Research PTC-100  
 Protocol:

Template: 10 ng  
 Primer: each 375 nM  
 dNTPs: each 200 uM  
 Tag Polymerase: 0.034 units/ul  
 Total Vol: 10 ul

Buffer:  
 MgCl2: 1.5 mM  
 KCl: 50 mM  
 Tris-HCl: 10 mM  
 pH: 8.3

FEATURES  
 source 1..436  
 Location/Qualifiers  
 1..436  
 /organism="Danio rerio"  
 /strain="AB"  
 /note="Vector: m13mp19 with added BstXI site; V-type:  
 Phase: Genomic DNA from a single adult zebrafish of AB  
 strain was digested with AluI, Cae8I, HaeIII, NlaVI, or  
 RsaI. Fragments in the range of 250-300 bp were gel  
 purified and a BstXI linker was added. The fragments were  
 cloned into a modified M13mp19 vector and transformed  
 into E. coli DH5alpha. Microsatellite sequences were  
 screened with labeled d(CA)15 and d(GT)15 oligonucleotide  
 probes."  
 /db\_xref="taxon:7955"  
 /clone="225840"  
 /clone\_1lb="zebrafish AB"  
 /sex="F"  
 /dev\_stage="Adult"  
 /lab\_host="DH5alphaF'10"  
 21..265  
 21..40  
 complement(246..265)  
 BASE COUNT 122 a 89 c 114 g 110 t 1 others

ORIGIN

Query Match 73.7% Score 14; DB 34; Length 436;  
 Best Local Similarity 100.0%; Pred. No. 5.42e+01;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 292 CATTATGAACGAT 305  
 |||||  
 Oy 4 CATTATGAACGAT 17

RESULT 8 AF073334 1515 bp mRNA INV 07-OCT-1998  
 LOCUS 225840.1 zebrafish AB Danio rerio STS genomic clone 225840 5',  
 DEFINITION sequence tagged site.  
 ACCESSION AF073334



REFERENCE	AUTHORS	TITLE	JOURNAL
1	Fan, J. and Brindley, P. J.	Retrotransposable elements in the <i>Schistosoma japonicum</i> genome	
2	(bases 1 to 1515)		
3	(11) Tada, I., Kojima, S. and Tsuji, M. (Eds.);	PROCEEDINGS OF THE 9TH INTERNATIONAL CONGRESS OF PARASITOLOGY	
4	(ICOPA IX): 821-825;		
5	Monduzzi Editore, Bologna, Italy (1998)		
6	2 (bases 1 to 1515)		
7	Fan, J. and Brindley, P. J.	Direct Submission	
8	Submitted (22-JUN-1998)	Molecular Parasitology Unit, Australia	
9	Centre for International and Tropical Health & Nutrition, Queensland		
10	Institute of Medical Research, P.O. Royal Brisbane Hospital, 300		
11	Herston Rd, Herston, Brisbane, Queensland 4029, Australia		
12	Location/Qualifiers		
13	1. .1515		
14	/organism="Schistosoma japonicum"		
15	/strain="Chinese"		
16	/db_xref="taxon:6182"		
17	/tissue_type="miracidium"		
18	/dev_stage="egg"		
19	<1. .1515		
20	/gene="RT"		
21	<1. .249		
22	/gene="RT"		
23	/note="Pao-like pol protein; truncated relative to other		
24	reverse transcriptases; similarity to reverse		
25	transcriptases extends beyond the stop codon"		
26	/codon_start=1		
27	/product="reverse transcriptase 'rtao'"		
28	/protein_id="AAC62956.1"		
29	/db_xref="PID:93702850"		
30	/db_xref="GI:3702850"		
31	/translation="ELTPAVLARIQSGLQELNREFSEVVLGTDSKIMUNYIRNESS		
32	RFAFVVGISAIHSHTFMQRYMPSKQNVANYPRG"		
33	BASE COUNT	460 a	250 c
34	ORIGIN	460 a	250 c
35	Query Match	73.7%	Score 14; DB 22; Length 1515;
36	Best Local Similarity 100.0%;		Pred. No. 5.42e+01;
37	Matches 14; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
38	DB	689	CTTCATTAATGAC 702
39	QY	1	CTTCATTAATGAC 14
40	RESULT	9	
41	LOCUS	DDU96916	1608 bp DNA
42	DEFINITION	Dictyostelium discoideum sks1 multidrug resistance protein homolog	18-MAY-1997
43	ACCESSION	U96916	
44	VERSION	92104756	
45	KEYWORDS	U96916.1	GI:2104756
46	SOURCE		
47	ORGANISM	Dictyostelium discoideum.	
48	REFERENCE	Dictyostelium discoideum	
49	AUTHORS	Eukaryotes: mitochondrial eukaryotes; Dictyostellida;	
50	TITLE	Dictyostelium.	
51	JOURNAL	1 (bases 1 to 1608)	
52	REFERENCE	De Maria, A.C. and Gomes, S.L.	
53	AUTHORS	Characterization of the sks1 multidrug resistance gene homolog in	
54	TITLE	Dictyostelium discoideum	
55	JOURNAL	Unpublished	
56	REFERENCE	2 (bases 1 to 1608)	
57	AUTHORS	De Maria, A.C. and Gomes, S.L.	

```

TITLE
JOURNAL Submitted (09-APR-1997) Departamento de Bioquímica, Universidade de
Sao Paulo, Av. Prof. Lineu Prestes, 748, sala 1207, Sao Paulo, SP
05508-900, Brasil
FEATURES
source
1..1608
/location/Qualifiers
/journal="Dictyostelium discoideum"
/db_xref="taxon:44689"
join(<260..392,490..749,863..1198,1315..>1506)
/product="sksl multidrug resistance protein homolog"
join(<260..392,490..749,863..1198,1315..1506)
/codon_start=1
/product="sksl multidrug resistance protein homolog"
/protein_id="AAB57823.1"
/db_xref="PID:g2104757"
/db_xref="GI:2104757"
/translation="MNRSLMSLGLREGLEGKRTDAPLPPTAETIHISLALKMLQOH
ARAGVPLEVMGLMIGELIDETIVIVFAMPQSTSVSEALDPVPTKMLDKOTQ
GRDIVIGWYHSHPGFCMLSVYNVNOSEPOLQSAVAVVDPDLSVSGKYIDAF
RTTSTPTAPROTTSNGLHLOPSIALIHGSRVYTAINYRKRELQKMLNLFH
KKRTLEGIDYKDETHQSNSEKQINMLLELTQYQKSIQEDKIEPKREYSAVGKLD
PKRRLLSDVITLVANNVRLVTLWLDIVT"
BASE COUNT 662 a 166 c 177 g 603 t
ORIGIN
Query Match 73.7% Score 14; DB 21; Length 1608;
Best Local Similarity 100.0%; Pred. No. 5,42e+01;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 264 ATCGTTCATTATG 277
|||||
|17 ATCGTTCATTATG 4
RESULT 10
LOCUS D89250 1724 bp mRNA PLN 13-MAR-1998
DEFINITION Schizosaccharomyces pombe mRNA, partial cds, clone: SY 1604.
ACCESSION D89250
NID g1749707
VERSION D89250.1 GI:1749707
KEYWORDS
SOURCE
ORGANISM
Schizosaccharomyces pombe (strain:PR745) cDNA to mRNA,
clone:lib:library of H. Nojima clone:SY1604.
Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Archiascomycetes;
Schizosaccharomycetales; Schizosaccharomycetales;
Schizosaccharomycetes.
1 (bases 1 to 1724)
Yoshioaka,S.
Direct Submission
Submitted (15-NOV-1996) to the DDBJ/EMBL/GenBank databases. Sachio
Yoshioaka, Tsukita Cell Axis Project ERATO JST, Kyoto Research Park;
17 Chidodaji Minamimachi, Shimokyo-ku, Kyoto, Kyoto 600, Japan
(E-mail:yoshio@cell.tsukita.jst.go.jp, Tel:+81-75-315-7913,
Fax:+81-75-315-6420)
2 (sites)
REFERENCE
AUTHORS Yoshioaka,S., Kato,K., Nakai,K., Okayama,H. and Nojima,H.
TITLE Identification of open reading frames in Schizosaccharomyces pombe
cDNAs
JOURNAL DDB Res. 4 (6), 363-369 (1997)
MEDLINE 9816272
FEATURES
source
Location/Qualifiers
1..1724
/journal="Schizosaccharomyces pombe"
/strain="PR745"
/db_xref="taxon:4896"
/codon_start=1
/codon_end=1724
/codon_id="SY1604"
/codon_lib="library of H. Nojima"
<1..1557
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA1911.1"
/db_xref="PID:d1014608"

```



/db\_xref="PID:g1749708"  
/db\_xref="GI:1749708"  
/translation="LIMMEEMROVNDANTELLRVNATLSOLKNTDMITVEOTSOIK  
EKRIIASLTKATKILNSASSIESORSEKSRSDSMRPOKEMVLMDGKHK  
DLKLTVEYAGRTFTTLKLLKNDVYQAKLDHFLAKLFIISOMKSNLCOESVA  
LVRVGVHTLEVYVOKTSSANERPDIGCOIMRSMLEFIISRLNIPRLSLFPP  
NLMNOCKACLOYESHLSKIFEFVVRVNNFLPLVQAMITGTINTDADENKRLRI  
KFEKPKYKIDFLHVNKVDSCQATVAVELYNALIRSIFFINVEATNSLFI  
GSKRGKINISYHVLKDWCSGPEAYQLDELQTSKILQFVDDPNVAVRDF  
YALNFOIKTLHRYDYADYEAHVPKKMSLSKNIAEGINQREOLTYEVLRIOD  
SFEPSLEKIKIPDCONVYLRNRIMDLAAESVSEGLITVGLTLMVRTATKTI  
KYOMVRFKN"

BASE COUNT 560 a 238 c 331 g 535 t

Query Match 73.7%; Score 14; DB 27; Length 1724;  
Best Local Similarity 100.0%; Pred. No. 5.42e+01;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 282 ATTAATGACGATG 295  
|||||  
5 ATTAATGACGATG 18

RESULT 11  
LOCUS MUSTPX101 1864 bp DNA ROD 11-JAN-1993  
DEFINITION Mouse testis-specific protein (TPX-1) gene, exon 1.  
ACCESSION M88617  
NID G202115  
VERSION M88617.1 GI:202115  
KEYWORDS testis-specific protein.  
SEGMENT 1 of 10  
SOURCE Mus musculus (strain BALB/cJ, sub-species domesticus) adult testis DNA.  
ORGANISM Mus musculus  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;  
Murinae; Mus.  
REFERENCE 1 (bases 1 to 1864)  
AUTHORS Mizuki, N., Sarapata, D. E., Garcia-Sanz, J. A. and Kasahara, M.  
TITLE The mouse male germ cell-specific gene Tpx-1: Molecular structure, mode of expression in spermatogenesis, and sequence similarity to two non-mammalian genes  
JOURNAL Mamm. Genome 3, 274-280 (1992)  
MEDLINE 92345716  
FEATURES  
source location/Qualifiers  
1. 1964  
/organism="Mus musculus"  
/strain="BALB/cJ"  
/sub-species="domesticus"  
/db\_xref="taxon:10090"  
/db\_stage="adult"  
/tissue\_type="testis"  
/map="Chromosome 17"  
1. 1964  
/gene="TPX-1"  
/number=1

BASE COUNT 589 a 377 c 397 g 601 t

Query Match 73.7%; Score 14; DB 32; Length 1964;  
Best Local Similarity 100.0%; Pred. No. 5.42e+01;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1022 GTTCATTAAATGAG 1035  
|||||  
14 GTTCATTAAATGAG 1

RESULT 12  
LOCUS ATUA6698 2316 bp mRNA PLN 22-MAR-1996  
DEFINITION Arabidopsis thaliana putative receptor serine/threonine kinase PR5K (PR5K) mRNA, complete cds.

ACCESSION U48698  
NID G1235679  
VERSION U48698.1 GI:1235679  
KEYWORDS  
SOURCE thale cress strain-Columbia.  
ORGANISM Arabidopsis thaliana  
Eukaryotes; mitochondrial eukaryotes; Viridiplantae;  
Charophyta; Embryophyta group; Embryophyta; Magnoliophyta;  
Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 2316)  
AUTHORS Wang, X., Zafian, P., Choudhary, M. and Lawton, M.  
TITLE The PR5K receptor protein kinase from Arabidopsis thaliana is structurally related to a family of plant defense proteins  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (5), 2598-2602 (1996)  
MEDLINE 96197359  
REFERENCE 2 (bases 1 to 2316)  
AUTHORS Lawton, M. A., Wang, X. Q. and Zafian, P.  
TITLE Direct Submission  
JOURNAL Submitted (06-FEB-1996) Michael A. Lawton, Agbiotech Center, Rutgers University, Foran Hall, Cook College, New Brunswick, NJ 08904, USA

FEATURES  
source location/Qualifiers  
1. 2316  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
115. 2112  
/chromosome="5"  
/gene="PR5K"  
115. 2112  
/gene="PR5K"  
/note="putative receptor serine/threonine kinase; displays serine/threonine protein kinase activity in vitro; extracellular domain similar to PR5 pathogens-related proteins"  
/codon\_start=1  
/product="receptor serine/threonine kinase PR5K"  
/protein\_id="AAC49208.1"  
/db\_xref="PID:g1235680"  
/db\_xref="GI:1235680"  
/translation="MVEGSLSMFLVSHFFVSGYMSNFTIENKCDYVMPGLM  
TIVASLPNNGFSLKGESEVYINPSPMSGLWRSFSTSTENSCATDCCSGKIE  
CSDSGARPTTLIDTFLDADGDFDVSVDGYNPLVVPQRLSGRCSVNGCV  
NLNKTCPSELKVGSSNKEHPICMNAACQFGEPCCEYEGKPAKCOPTLYSTNK  
NECPPLAYSAVDNENNTFRCSNPNVITFCPDIDISMSQSPKEKNGTOKSSMKL  
LTVGSAALIMLIVVIVITFRKNRNSMNNOANEAVALMKRYATIRKMTNSPA  
HTVGSGFGVYVYGRKLADSGRDYAVVILKYSBENGSEFNEVASMRTSVNIVSLG  
FCEKNNRATITTFMPNGSLDKTISANMSTKMEERLYDVAVIGKLELHNRVTR  
IVHFDIKPQNLMDENLCPKISDFGLAKLCKNESLISMLHMGITGGLAPENFSKNF  
GAVSHKSDVYSYGMVVLMEGAKNIKEVERSENNNSMFEPPVWYRDFEKGELTRIFG  
NSTIEEEKFAKTLVLAWLCIOMNSDPRPMKYVTEMLGNEALQVPPNPPLFSPPE  
ETVPTLLESDDTSTFNPSPHFRGLLNSSEVLYLQHSRS"

BASE COUNT 689 a 443 c 527 g 657 t

Query Match 73.7%; Score 14; DB 28; Length 2316;  
Best Local Similarity 100.0%; Pred. No. 5.42e+01;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1227 GTTCATTAAATGAG 1240  
|||||  
14 GTTCATTAAATGAG 1

RESULT 13  
LOCUS ANNAD2 2609 bp DNA PLN 25-MAR-1999  
DEFINITION A.nidus mitochondrial nad2 gene.  
ACCESSION Y07910  
NID G1546880  
VERSION Y07910.1 GI:1546880  
KEYWORDS nad2 gene; NADH dehydrogenase.  
SOURCE bird's-nest fern.  
ORGANISM Mitochondrion Asplenium nidus



```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Euphyllophytes; Filicophyta; Filicopsida; Filicales; Aspleniaceae;
Asplenium.
REFERENCE      1 (bases 1 to 2609)
AUTHORS       Malek,O. and Knopp,V.
TITLE          Trans-splicing group II introns in plant mitochondria: the complete
JOURNAL        RNA 4 (12), 1599-1609 (1998)
MEDLINE        99063402
REFERENCE      2 (bases 1 to 2609)
AUTHORS       Malek,O.
TITLE          Direct Submission
JOURNAL        Submitted (11-SEP-1996) O. Malek, Inst.f. Genbiologische Forschung,
               Ihnestr. 63, D-14195 Berlin, FRG
FEATURES       Location/Qualifiers
               1..2609
               /organism="Asplenium nidus"
               /mitochondrion
               /db_xref="taxon:29642"
               join(<1..119,2465..>2609)
               /gene="nad2"
               /note="subunit 2"
               /product="NADH dehydrogenase"
               1..2609
               /gene="nad2"
               <1..119
               /gene="nad2"
               /number=1
               2
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               4
               /gene="nad2"
               /note="U to C RNA editing"
               11
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               19
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               55
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               80
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               113
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               117
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               120..2464
               /gene="nad2"
               /note="group II"
               /number=1
               /evidence=experimental
               2465..>2609
               /gene="nad2"
               /number=2
               2521
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               2608
               /gene="nad2"

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Euphyllophytes; Filicophyta; Filicopsida; Filicales; Aspleniaceae;
Asplenium.
REFERENCE      1 (bases 1 to 2609)
AUTHORS       Malek,O. and Knopp,V.
TITLE          Trans-splicing group II introns in plant mitochondria: the complete
JOURNAL        RNA 4 (12), 1599-1609 (1998)
MEDLINE        99063402
REFERENCE      2 (bases 1 to 2609)
AUTHORS       Malek,O.
TITLE          Direct Submission
JOURNAL        Submitted (11-SEP-1996) O. Malek, Inst.f. Genbiologische Forschung,
               Ihnestr. 63, D-14195 Berlin, FRG
FEATURES       Location/Qualifiers
               1..2609
               /organism="Asplenium nidus"
               /mitochondrion
               /db_xref="taxon:29642"
               join(<1..119,2465..>2609)
               /gene="nad2"
               /note="subunit 2"
               /product="NADH dehydrogenase"
               1..2609
               /gene="nad2"
               <1..119
               /gene="nad2"
               /number=1
               2
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               4
               /gene="nad2"
               /note="U to C RNA editing"
               11
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               19
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               55
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               80
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               113
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               117
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               120..2464
               /gene="nad2"
               /note="group II"
               /number=1
               /evidence=experimental
               2465..>2609
               /gene="nad2"
               /number=2
               2521
               /gene="nad2"
               /note="C to U RNA editing"
               /evidence=experimental
               2608
               /gene="nad2"

Query Match      73.7%; Score 14; DB 28; Length 2609;
Best Local Similarity 100.0%; Pred. No. 5.42e+01;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
ORIGIN
BASE COUNT      698 a      662 c      638 g      611 t
misc_feature     /note="C to U RNA editing"
                 /evidence=experimental
                 2609
                 /gene="nad2"
                 /note="C to U RNA editing"
                 /evidence=experimental

RESULT 14
LOCUS      BACHSMR      2892 bp      DNA
DEFINITION Bacillus subtilis ISB8 modification methyltransferase (hsmBI) gene,
complete cds, restriction endonuclease (hsrBI) gene, complete cds.
ACCESSION  G143053
VERSION    L01541.1 GI:143053
KEYWORDS   hsmBI gene; hsrBI gene; modification methyltransferase; restriction
           endonuclease.
SOURCE     Bacillus subtilis (strain ISB8) DNA.
ORGANISM   Bacillus subtilis
            Eubacteria; Firmicutes; Low G+C gram-positive bacteria;
            Bacillaceae; Bacillus.
REFERENCE  1 (bases 1 to 2892)
AUTHORS   Xu,G.L., Kapfer,W., Walter,J. and Trautner,T.A.
TITLE     Bsubl--an isospecific restriction and modification system of PstI:
           characterization of the BsubI genes and enzymes
JOURNAL    Nucleic Acids Res. 20 (24), 6517-6523 (1992)
MEDLINE    93126092
FEATURES   Location/Qualifiers
           1..2892
           /organism="Bacillus subtilis"
           /db_xref="taxon:1423"
           185..190
           /gene="hsmBI"
           198..>1733
           /gene="hsmBI"
           214..220
           /gene="hsmBI"
           228..1733
           /gene="hsmBI"
           228..1733
           /gene="hsmBI"
           /codon_start=1
           /transl_table=11
           /product="modification methyltransferase"
           /protein_id="AA18169.1"
           /db_xref="PID:9143054"
           /db_xref="GI:143054"
           /translation="MTOILETVKSRLLVNPFLNKKSLGQFFPPSSISIFMACLSE
           DKINNAKVLDAAGAGIGSLTSAFIRLTISENTGKADHILLETDEMPELSTLAFKD
           YIETNSQITITDFITMANYSLLDESSLAKDKORTAHILNPPYKIKISNKKRLR
           KAGITVNLVSFAFVALTVLMSDGEIVFIIPRSFQNPYFRHFRQHLNKTSTKHM
           LFEERDKAFKDEVLQENVISLKEKGVQEDVKISISTDSFSVRSYRYPEKIVOP
           .NDIEKFIHINTNEETLIEKHPNVCYSLEELNIYVSTGPVDFRVKENTLRMPGEGTV
           PLEFVNHVGTSLSEYPKMKKPNATIRNEKVKMLYPGHVVVVKRFSKKKKRIVA
           GVLPPESVNDPVVGFENGLNLYHKVKSISKEVAGLYAINTSPVDKPYRIFNCHQ
           VNAIDLRTMKFPSPRDLISLGKVIENIENYGOVEIDSKLELLNDNGNA"
           1697..1701
           /gene="hsmBI"
           1714..2664
           /gene="hsrBI"
           1714..2664
           /gene="hsrBI"
           /codon_start=1

```







```

/db_xref="GI:4007778"
/translation="MVESKIKLLEQVQADPNRSIQALGRDLREKHYHDPANLTDI
ASSGASVPLROGSLIYLQRYIVHHSPLFEQFDGPIDENVKKHVRFTLLHLVSLD
NETLIKAVAYVASLIANDVDPDENPEVPAVLHLQSSNENSINSLDLVLELDES
VEBOEFTIAPOLASILVOFTSAPDSMRLOARGICLFRSCLELLEIYETKELPSK
RVFLAEQIIPPMNMFHKEFVSLVDDRVTLLPDSGICFCINGEIAMTITKRLPEPSK
LTPYVTFVELWMNITEKLDPIREVVSGLDSDAGDKYPIRYVELLELVVAL
QSKFQVNLVSNTPVPPPCIPLLVOYIOLPKHQLVSEVSDVSEYIANEFSMDPAS
DITVGAISVLSAEFEHTTLPQQSLREMSATYILNNIMTYOELIYACSDVAS
DDTDDYLDPIYEAIAKVIDSDAPILLSEPFLEIGYSESTVVASOFOIIMNNV
NALOVDTVOYAAKAIERFGSGVKRPLISLOPMILEVLSQYASKSSDEALVLEAI
SSAYKICAKAEICGNSYPIPLFNUNANADRYICGILETFEDITHAANYSMCE
ITLPELLOVLMQEDPIMNIGATILSCILIRGAPSPLEPGVGLPVPYKITQHSBG
TELLQSOELLKLEKDPQLLEISGSSGFQYILFLHLQLDKESDSDGACFLVGP
ILLELADHASQMDVLOSILSCIKRAIAEPRIOSIIVYFAKLIYDSLSGMHFLT
SSLNEOGLTAFEVLMYWCNDFVSNFKNISIIICIMTKIYFSDSLDSYGVNGE
LISHSNRIITRSOSKIHPEEYSYVSGVKIRLLISEEVSRSKAIYEVSDGADM
DDGPISAFTEGLSANDFSGVNSDEEDNTLOLYLLEFFEKAKNSLNHINVEVRL
PQEQDALVQIKER"
misc_feature
complement(3121..3133)
/gene="SPAC22H10.03c"
/note="splice branch and acceptor sequence, ctacattttag"
complement(3154..3159)
/gene="SPAC22H10.03c"
/note="splice donor sequence, gtgggt"
complement(3213..3230)
/gene="SPAC22H10.03c"
/note="splice branch and acceptor sequence,
ctacgaattgtccaag"
complement(3264..3269)
/gene="SPAC22H10.03c"
/note="splice donor sequence, gtacga"
complement(3441..3454)
/gene="SPAC22H10.03c"
/note="splice branch and acceptor sequence,
ttaacaaactag"
complement(3477..3482)
/gene="SPAC22H10.03c"
/note="splice donor sequence, gtagt"
complement(3602..3613)
/gene="SPAC22H10.03c"
/note="splice branch and acceptor sequence, ctacactgaag"
complement(3643..3648)
/gene="SPAC22H10.03c"
/note="splice donor sequence, gtagt"
complement(3819..3834)
/gene="SPAC22H10.03c"
/note="splice branch and acceptor sequence,
ctgaattatttag"
complement(3853..3858)
/gene="SPAC22H10.03c"
/note="splice donor sequence, gtagt"
complement(5917..5943)
/gene="SPAC22H10.03c"
/note="PS00215 Mitochondrial energy transfer proteins
signature"
complement(6081..6099)
/gene="SPAC22H10.03c"
/note="splice branch and acceptor sequence,
ctaactattttag"
complement(6131..6136)
/gene="SPAC22H10.03c"
/note="splice donor sequence, gtagt"
7212..8313
/gene="SPAC22H10.04"
join(7212..7372,7439..7509,7622..8313)
/gene="SPAC22H10.04"
/note="SPAC22H10.04, len: 307, similar to PIR:S42558
phosphoprotein phosphatase (58.0% identity in 307 aa
overlap): PS00125 Serine/threonine specific protein
phosphatases signature"
/codon_start=1
/product="putative phosphoprotein phosphatase"

```

```

/protein_id="CAA93605.1"
/db_xref="PID:e223751"
/db_xref="PID:g1204185"
/db_xref="GI:1204185"
/db_xref="SWISS-PROT:Q10298"

```

Note: remainder of annotations omitted.

```

Query Match      73.7%  Score 14; DB 27; Length 27227;
Best Local Similarity 100.0%; Pred. No. 5.42e+01;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db      8409 CATCGTATTAAAT 8422
Cp      18 CATCGTATTAAAT 5

```

Search completed: Sat Nov 27 15:08:51 1999  
Job time : 135 secs.







00GNTWVHVPSPRLNANVKAVERKFGAIVPMFOALSECTPYDINOMLNCZDH  
OCALQIVKEIINEEAQMDIOHPPAGIPAGLPRGSDVAGTSTVOEDLEMYT  
ANPRVDGAIYRRMIIIGLQCKVKMYNPISIDIKGPREPKDYDRFYRTIRAEQA  
SGEVKO"

BASE COUNT 245 a 141 c 161 g 125 t  
ORIGIN approximately 304 bp downstream from the start of gag

Query Match 63.2%; Score 12; DB 3; Length 672;  
Best Local Similarity 100.0%; Pred. No. 7.33e-03;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 335 tccatgaag 346  
|||||  
CP 12 TCATTAATGAG 1

RESULT 2 HIV27924A1 781 bp ds-DNA VRL 23-AUG-1994  
LOCUS Human immunodeficiency virus type 2 (HIV-2) p16/p28 gag sequence,  
DEFINITION sample 7924A, clone 1, partial cds.  
ACCESSION L33081  
SOURCE Human immunodeficiency virus type 2 PCR-amplified DNA recovered from  
a 48 year old heterosexual male originally from urban Guinea Bissau  
who showed HIV dual seroreactivity.  
ORGANISM Human immunodeficiency virus type 2  
Virusidae; ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
1 (bases 1 to 781)

REFERENCE Gao,F., Yue,L., Robertson,D.L., Hill,S.C., Hul,H., Bisgar,R.J.,  
Neugebue,A.E., Whelan,T.M., Ho,D.D., Shaw,G.M., Sharp,P.M. and  
Hahn,B.H.

TITLE Genetic diversity of human immunodeficiency virus type 2: Evidence  
for five distinct sequence subtypes with differences in virus  
biology  
J. Virol. 68, 7433-7447 (1994)

JOURNAL full staff\_review  
STANDARD This sequence was obtained as part of a study designed to determine  
COMMENT the in vivo extent of HIV-2 genetic and biological variation. Up  
to now, HIV-2 sequences have been derived almost solely from  
cultured isolates. Since cultivation may select against viruses  
which can not grow in a particular cell type, Gao et al. argue that  
a greater variation may be elucidated by sequencing virus obtained  
from uncultured PBMCs. With this aim in mind, peripheral blood  
samples were collected from 12 HIV-2 seropositive patients (2238,  
60415K, 60667K, 7312A, 7810A, 7924A, FA, F0784, FT, JA, ON, and PA)  
from six different West African countries, living in both urban and  
rural areas. Uncultured PBMC DNA was PCR amplified, and the  
recombinant clones were manually sequenced. Another portion of the  
cells was used for virus isolation.  
Patient 7924 was a heterosexual 48-year old teacher from Guinea  
Bissau, who was living in the Washington, D.C. area. Originally he  
lived in an urban area in West Africa before traveling extensively  
and coming to the United States. After seeking medical care for a  
STD, patient 7924 was identified as HIV-2 positive. Subsequent  
Western blot analysis of peripheral blood cells showed HIV dual  
reactivity. Virus isolation was successfully performed at the  
University of Alabama at Birmingham. Sample 7924A clusters with  
the subtype A HIV-2 viruses.  
NCBI gi: 532145

FEATURES  
source Location/Qualifiers  
1..781  
/organism="Human immunodeficiency virus type 2"  
/proviral  
/sequenced\_mol="DNA"  
/rissue\_type="PBMC"  
<1..781  
/partial  
/gene="gag"  
/note="NCBI gi: 532146"  
/codon\_start=2  
/translation="ANELDRFLAESLSKRCORRISVDPLPYTGSENGKSLFNT  
VAVVCIHAEKVDEAKKIIVQRHLVAETGTAEMPTSTPAPPSKGGNIYVQS  
IGGNTHTVPLSPRTLNANVKAVERKFGAIVPMFOALSECTPYDINOMLNCZDHQ

AAMQIREINEVAADWDVAPHPGIPPLAGLREPRGSDIAGTSTVEEQIOWMFRQO  
NPEVGNIRRMVIGLQCKVKMYNPISIDIKGPREPKDYDRFYRTIRAEQA  
SGEVKO"

BASE COUNT 279 a 157 c 195 g 150 t  
ORIGIN 111 bp downstream from the start of gag

Query Match 63.2%; Score 12; DB 3; Length 781;  
Best Local Similarity 100.0%; Pred. No. 7.33e-03;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 507 tccatgaag 518  
|||||  
CP 12 TCATTAATGAG 1

RESULT 3 SIVAGMT49 1773 bp ds-DNA VRL 09-JUN-1993  
LOCUS Simian immunodeficiency virus, isolate TAN-49, complete gag cds.  
DEFINITION L19253  
ACCESSION  
SOURCE Simian immunodeficiency virus proviral DNA, isolate TAN-49 from a  
Ugandan tantalus monkey (Cecopithecus tantalus).  
ORGANISM Simian immunodeficiency virus  
Virusidae; ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
1 (bases 1 to 1773)

REFERENCE Hirsch,V.M., McGinn,C., Dapolito,G., Goldstein,S., Ogen-Odoi,A.,  
Biriyawho,B., Lakwo,T. and Johnson,P.R.  
Identification of a new subgroup of SIVagm in tantalus monkeys  
Virology (1993) In press

TITLE full staff\_review  
JOURNAL kindly submitted prior to publication by Dr. Vanessa Hirsch,  
STANDARD Immunodeficiency Viruses Section, NIAID/NIH, Twindbrook II,  
COMMENT Rockville MD 20852. SIVagm-tan strains cluster with other  
previously characterized agm strains taken from vervets and  
grivets. See related tantalus isolate sequences 17, 27, 40, and  
9 (accession numbers L19250-L19252, L19254).

FEATURES  
exon Location/Qualifiers  
155..1773  
/gene="gag"  
1..1773  
/organism="Simian immunodeficiency virus"  
/cell\_line="CEMS"  
/haplo\_type="na"  
/proviral  
/sequenced\_mol="DNA"

BASE COUNT 594 a 375 c 480 g 324 t  
ORIGIN 193 nt upstream from the gag cds start.

Query Match 63.2%; Score 12; DB 3; Length 1773;  
Best Local Similarity 100.0%; Pred. No. 7.33e-03;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 828 tccatgaag 839  
|||||  
CP 12 TCATTAATGAG 1

RESULT 4 SIVAGMTYO 9170 bp ss-RNA VRL 01-MAR-1989  
LOCUS Simian (African green monkey) immunodeficiency virus, isolate  
DEFINITION TYO-1, complete genome.  
ACCESSION X07805  
SOURCE Simian (African green monkey) immunodeficiency virus, extra-  
chromosomal closed-circular DNA, TYO-1 clone lambda-SAH12.  
REFERENCE 1 (bases 1 to 9170)  
AUTHORS Fukasawa,M., Mura,T., Hasegawa,A., Morikawa,S., Tsujimoto,H.,  
Miki,K., Kitamura,T. and Hayami,M.  
Sequence of simian immunodeficiency virus from African green  
monkey, a new member of HIV/SIV group  
Nature 333, 457-461 (1988)  
JOURNAL full staff\_review  
STANDARD  
COMMENT

This sequence was obtained from a lambda clone derived from an  
African green monkey imported from Kenya.







```

      exon          6135..6195
                    /number=2
                    /gene="rev"
CDS              6202..8766
                    /gene="env"
                    /codon_start=6202
      exon          8368..8449
                    /number=3
                    /gene="lat"
                    /number=3
      exon          8368..8750
                    /number=3
                    /gene="rev"
                    /gene="rev"
      CDS          8600..9271
                    /gene="nef"
                    /codon_start=8600
      LTR          8937..9623
                    /note="3', long terminal repeat"
      repeat_unit  9407..9623
                    /type="R repeat 3' copy"
      polyA_signal 9506..9511
      BASE COUNT  3243 a 1807 c 2470 g 2103 t
      ORIGIN      5' terminus of 5' LTR.

Query Match      63.2%; Score 12; DB 3; Length 9623;
Best Local Similarity 100.0%; Pred. No. 7.33e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1513 tcatatgaag 1524
|||||
12 TCATTATGAG 1

RESULT 6
LOCUS      SIY2010G      672 bp DNA      VRL      24-AUG-1994
DEFINITION Simian immunodeficiency virus (SIYbab) partial gag sequence.
ACCESSION  U10899
SOURCE      Simian immunodeficiency virus from PCR-amplified DNA extracted from
            uncultured leukocytes derived from an asymptomatic pregnant female
            yellow baboon from Tanzania.
ORGANISM    Simian immunodeficiency virus
            Vira; Viruses; ssRNA enveloped viruses; Retroviridae; Lentivirus;
            Primate immunodeficiency viruses.
REFERENCE   1 (bases 1 to 672)
            Jin,M.J., Rogers,J., Phillips-Conroy,J.E., Allan,J.S.,
            Desrosiers,R.C., Shaw,G.M., Sharp,P.M. and Hahn,B.H.
            Infection of a yellow baboon with SIV from African green monkeys:
            evidence for cross-species transmission in the wild
            J. Virol. (in press, 1994)
TITLE       J. Virol. (in press, 1994)
JOURNAL     Full staff review
STANDARD   Two wild yellow baboons from Tanzania were identified in a previous
COMMENT     study with strong SIV(agm) seroreactivity. They lived in a
            national park where yellow baboons and African green monkeys share
            the same habitat. In this study, molecular characterization was
            performed for virus derived from one of the SIVagm seroreactive
            yellow baboons (animal 2010) and from two vervet African green
            monkeys (animals ver266 and ver385). With this aim in mind,
            uncultured leukocyte DNA was PCR-amplified, cloned and sequenced
            over a 672 bp region of gag and a 906 bp region of env for animal
            2010 (only the env region was analyzed for isolates ver266 and
            ver385). Jin et al. note that the viral strain derived from
            the yellow baboon (animal 2010) clusters with SIV strains of the
            vervet subtype, including ver266 and ver385, and believe that
            this is the first evidence for simian-to-simian cross-species
            transmission of SIV in the wild.
FEATURES
  Source      location/Qualifiers
  1..672
  /clone="gagbab2010"
  /strain="SIYbab2010"
  /organism="Simian immunodeficiency virus"
  /specific_host="Papio cynocephalus-hamadryas"
  <1..>672
  /gene="gag"

```

```

      /product="gag protein"
      /translation="VATVROCCHLVKKESAETSSGRKNDSTATSSGQSNFPQA
      OGGNTWVHVPLEPTINAMKAVKEKFGAEIYVPALEGGCPYINOMLVNIGDH
      OGALQIKKEIINEEPAQMDIOHPPAKPLPAGOLRDRGSDVAGTSTVQOLEMIYT
      ANPRVDGAIIRRIIILGLOKCYKMPVSIIDIKQSPKPPADYVRFRTIRAEQA
      SEGVKQ"
BASE COUNT      245 a 141 c 161 g 125 t
ORIGIN          approximately 304 bp downstream from the start of gag

Query Match      57.9%; Score 11; DB 3; Length 672;
Best Local Similarity 100.0%; Pred. No. 1.33e-01;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 335 tcatatgaag 345
|||||
07 3 TCATTATGAA 13

RESULT 7
LOCUS      HIV27924A1      781 bp ds-DNA      VRL      23-AUG-1994
DEFINITION Human immunodeficiency virus type 2 (HIV-2) p16/p28 gag sequence,
            sample 7924A, clone 1, partial cds.
ACCESSION  L33081
SOURCE      Human immunodeficiency virus type 2 PCR-amplified DNA recovered from
            a 48 year old heterosexual male originally from urban Guinea Bissau
            who showed HIV dual seroreactivity.
ORGANISM    Human immunodeficiency virus type 2
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1 (bases 1 to 781)
            Gao,F., Yue,L., Robertson,D.L., Hill,S.C., Hui,H., Biggar,R.J.,
            Neeguaye,A.E., Whelan,T.M., Ho,D.D., Shaw,G.M., Sharp,P.M. and
            Hahn,B.H.
            Genetic diversity of human immunodeficiency virus type 2: Evidence
            for five distinct sequence subtypes with differences in virus
            biology
            J. Virol. 68, 7433-7447 (1994)
TITLE       Full staff review
JOURNAL     This sequence was obtained as part of a study designed to determine
STANDARD   the in vivo extent of HIV-2 genetic and biological variation. Up
COMMENT     to now, HIV-2 sequences have been derived almost solely from
            cultured isolates. Since cultivation may select against viruses
            which can not grow in a particular cell type, Gao et al. argue that
            a greater variation may be elucidated by sequencing virus obtained
            from uncultured PBMCs. With this aim in mind, peripheral blood
            samples were collected from 12 HIV-2 seropositive patients (2238,
            60415K, 60667K, 7312A, 7810A, 7924A, FR, FO784, FT, JA, ON, and PA)
            from six different West African countries, living in both urban and
            rural areas. Uncultured PBMC DNA was PCR amplified, and the
            recombinant clones were manually sequenced. Another portion of the
            cells was used for virus isolation.
            Patient 7924 was a heterosexual 48-year old teacher from Guinea
            Bissau, who was living in the Washington, D.C. area. Originally he
            lived in an urban area in West Africa before traveling extensively
            and coming to the United States. After seeking medical care for a
            STD, patient 7924 was identified as HIV-2 positive. Subsequent
            Western blot analysis of peripheral blood cells showed HIV dual
            reactivity. Virus isolation was successfully performed at the
            University of Alabama at Birmingham. Sample 7924A clusters with
            the subtype A HIV-2 viruses.
            NCBI gi: 532145
FEATURES
  Source      location/Qualifiers
  1..781
  /organism="Human immunodeficiency virus type 2"
  /proviral
  /sequenced_mol="DNA"
  /tissue_type="PBMC"
  <1..>781
  /partial
  /gene="gag"
  /note="NCBI gi: 532146"
  /codon_start=2

```



```

/translation="ANEDPRFGAESLSEKSCORITISVDPIVPTGSENLKSLFNT
VCVVCIAHEEKVKDTEBAKTIQRLVAETGAEMPTSRAPSGKGVYQVS
IGNATVHPLSPRTLNANWALVEKKEFGAEVPGFALSGCTPDINDMLCVGDH
AAAOIIRIINEVAADMDVAHPIGGLPAQREPGSDIAGTSTVEBOIOMFERAO
NVPVGNIRRMIOIGLCKRCVMYNTNILDVKGKESFQS"
BASE COUNT      279 a 195 g 150 t
ORIGIN          111 bp downstream from the start of gag

Query Match      57.9%; Score 11; DB 3; Length 781;
Best Local Similarity 100.0%; Pred. No. 1.33e-01;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 507 tcattatgaa 517
|||||
OY 3 TCATTATGAA 13

RESULT 8
LOCUS          HUMB2M1 1217 bp ds-DNA PRI 15-JUN-1989
DEFINITION    Human beta-2-microglobulin gene, exon 1.
ACCESSION     M17986
SEGMENT       1 of 2
SOURCE        Human DNA, clone pB2m13, and cDNA to mRNA, clone pB2m16.
ORGANISM      Homo sapiens
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
Eutheria; Primates; Haplorhini; Catarrhini; Homnidae.
REFERENCE     Gnessow,D., Rein,R., Ginjaar,I., Hochstenbach,F., Seemann,G.,
AUTHORS       Kottman,A. and Ploegh,H.L.
TITLE         The human beta-2-microglobulin gene: Primary structure and
              definition of the transcriptional unit
JOURNAL       J. Immunol. 139, 3132-3138 (1987)
STANDARD     Draft entry and computer-readable sequence for [1] kindly provided
COMMENT       by E.J.Baas, 03-DEC-1987.

FEATURES
  exon
    location/Qualifiers
      <883..949
      /number=1
      /note="beta-2-microglobulin precursor"
      /gene="B2M"
      /map="15q21-q22.2"
      prim_transcript 846..>1217
      sig_peptide      883..942
      /note="microglobulin mRNA and intron"
      /note="beta-2-microglobulin, signal peptide"
      /codon_start=1
      943..949
      /note="beta-2-microglobulin, mature peptide"
      /codon_start=1
      950..>1217
      /note="intron A"
      1..1217
      source
        /organism="Homo sapiens"
BASE COUNT      274 a 325 c 322 g 296 t
ORIGIN          383 bp upstream of EcolI site.

Query Match      57.9%; Score 11; DB 1; Length 1217;
Best Local Similarity 100.0%; Pred. No. 1.33e-01;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 32 catcgtcatt 42
|||||
Cp 18 CATCGTTCATT 8

RESULT 9
LOCUS          SIVAGMT49 1773 bp ds-DNA VRL 09-JUN-1993
DEFINITION    Simian immunodeficiency virus, isolate TAN-49, complete gag cds.
ACCESSION     L19253
SOURCE        Simian immunodeficiency virus proviral DNA, isolate TAN-49 from a
              Ugandan tantalus monkey (Cecopithecus tantalus).
ORGANISM      Simian immunodeficiency virus
              Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;

```

```

Retroviridae; Lentivirinae.
1 (bases 1 to 1773)
AUTHORS       Hirsch,V.M., McGowan,C., Dapolito,G., Goldstein,S., Ogen-Odoi,A.,
              Birwaaho,B., Lakwo,T. and Johnson,P.R.
TITLE         Identification of a new subgroup of SIVagm in tantalus monkeys
JOURNAL       Virology (1993) In press
STANDARD     full staff_review
COMMENT       kindly submitted prior to publication by Dr. Vanessa Hirsch,
              Immunodeficiency Viruses Section, NIAID/NIH, Twinbrook II,
              Rockville MD 20852. SIVagm-tan strains cluster with other
              previously characterized agm strains taken from vervets and
              givrets. See related tantalus isolate sequences 17, 27, 40, and
              9 (accession numbers L19250-L19252, L19254).
              Location/Qualifiers
                exon
                  /gene="gag"
                  1..1773
                  /organism="Simian immunodeficiency virus"
                  /cell_line="CEM5"
                  /haplotype="na"
                  /proviral
                  /sequenced_mol="DNA"
                  375 c 480 g 324 t
BASE COUNT      594 a 375 c 480 g 324 t
ORIGIN          193 nt upstream from the gag cds start.

Query Match      57.9%; Score 11; DB 3; Length 1773;
Best Local Similarity 100.0%; Pred. No. 1.33e-01;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 828 tcattatgaa 838
|||||
OY 3 TCATTATGAA 13

RESULT 10
LOCUS          SIVAGMTYO 9170 bp ss-RNA VRL 01-MAR-1989
DEFINITION    Simian (African green monkey) immunodeficiency virus, isolate
              TYO-1, complete genome.
ACCESSION     X07805
SOURCE        Simian (African green monkey) immunodeficiency virus, extra-
              chromosomal closed-circular DNA, TYO-1 clone lambda-SAH12.
REFERENCE     1 (bases 1 to 9170)
              Fukasawa,M., Miura,T., Hasegawa,A., Morikawa,S., Tsujimoto,H.,
              Miki,K., Kitamura,T. and Hayami,M.
              Sequence of simian immunodeficiency virus from African green
              monkey, a new member of HIV/SIV group
JOURNAL       Nature 333, 457-461 (1988)
STANDARD     full staff_review
COMMENT       This sequence was obtained from a lambda clone derived from an
              African green monkey imported from Kenya.

SIVAGMTYO appears to lack a vpr coding region. [1] suggests that
this may account for SIVAGMTYO's lack of pathogenicity. The env
cds contains an in-frame stop codon at positions 8077-8079, similar
to the in-frame stop codons found in the SIV macaque and HIV-2
RD5 envelope coding regions, except that SIVAGMTYO's stop codon
is found 96bp downstream from the 3' splice junction of tat and
rev, which is the location of the stop codon in the other
sequences.

[1] finds that SIVAGMTYO is approximately equally distantly related
to the HIV-1's and the HIV-2/SIV (macaque) group; over a
"conserved" stretch of the pol protein, TYO is only 87% similar to
AGM385 and 82% similar to AGM265, the latter two differing by 20%.
Location/Qualifiers
  LTR
    <1..219
    /note="5' LTR"
    <1634..4819
    /note="pol polypeptide (NH2-terminus uncertain; AA at
    1634)"
    /codon_start=1634
    join(5546..5763,7984..8068)

```



```

CDS      /note="cat protein"
         /codon_start=5546
         join(5703..5763,7984..8177)
         /note="rev protein"
         /codon_start=5703
prim_transcript 1..9170
         /note="genomic mRNA"
prim_transcript 1..9170
         /note="cat, rev, nef subgenomic mRNA"
repeat_region 1..117
         /note="R repeat 5' copy"
misc_binding 220..237
         /note="primer (Lys-tRNA) binding site"
CDS      432..1991
         /note="gag polyprotein"
         /codon_start=432
CDS      4755..5462
         /note="vif protein"
         /codon_start=4755
CDS      5236..5595
         /note="vpx protein"
         /codon_start=5236
exon     5546..5763
         /note="tat protein, exon 2 (first expressed exon)"
exon     5703..5763
         /note="rev protein, exon 2 (first expressed exon)"
intron   5764..7983
         /note="cat intron 2"
intron   5764..7983
         /note="rev intron 2"
intron   5764..7983
         /note="cat, rev, nef subgenomic mRNA intron 2"
CDS      5770..8367
         /pseudo
         /note="env polyprotein (in-frame stop codon at 8077)"
         /codon_start=5770
exon     7984..8068
         /note="cat protein, exon 3 (AA at 7985)"
exon     7984..8177
         /note="rev protein, exon 3 (AA at 7986)"
misc_feature 8077..8079
         /note="in-frame stop in env cds"
CDS      8201..8890
         /note="nef protein"
         /codon_start=8201
LTR      8547..>9170
         /note="3' LTR"
repeat_region 9054..9170
         /note="R repeat 3' copy"
misc_signal 9151..9156
         /note="mRNA polyadenylation signal"
BASE COUNT 3172 a 1720 c 2263 g 2015 t
ORIGIN    Cap site of genomic RNA.
Query Match 57.9%; Score 11; DB 3; Length 9170;
Best Local Similarity 100.0%; Pred. No. 1.33e-01;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 1066 tcattaatgaa 1076
      |||
QY 3 TCATTAATGAA 13

```

```

TITLE    A highly divergent proviral DNA clone of STV from a distinct
          species of African green monkey.
JOURNAL   Virology 182, 397-402 (1991)
STANDARD  full staff entry
REFERENCE 2 (bases 1 to 2438)
AUTHORS   Johnson,P.R., Fomsgaard,A., Allan,J., Gravel,M., London,W.T.,
          Olmstead,R.A. and Hirsch,V.M.
TITLE     Simian immunodeficiency viruses from African green monkeys display
          unusual genetic diversity
JOURNAL   J. Virol. 64, 1086-1092 (1990)
STANDARD  full staff entry
COMMENT   kindly submitted prior to publication and in a computer readable
          form by Phillip Johnson, NIAID/Georgetown University.
          The typical stop codon "tag" in rev (position 8574) is "cag"
          making rev in SIVgr1-1 (lambdaII) longer than rev in most
          other sequences. The gr1-1 sequence is the most divergent
          primate lentivirus reported to date.
          location/Qualifiers
FEATURES  join(5993..6195,8370..8750)
          /gene="tat"
          /codon_start=5993
          /partial
          join(6135..6195,8371..8750)
          /note="The normal stop codon 'tag' in rev (position 8574)
          is 'cag' making rev in SIVgr1-1 longer than rev in most
          other sequences. The gr1-1 sequence is the most divergent
          primate lentivirus reported to date."
          /gene="rev"
          /codon_start=6135
          1..689
          /note="5' long terminal repeat"
          1..587
          /rpt_type=R repeat 5' copy
          690..707
          /note="primer (Lys-tRNA) binding site"
          897..2438
          /gene="gag"
          /codon_start=897
          2141..5314
          /gene="pol"
          /codon_start=2141
          /partial
          5214..5873
          /gene="vif"
          /codon_start=5214
          5683..6039
          /gene="vpx"
          /codon_start=5683
          5993..6195
          /number=2
          /gene="tat"
          6135..6195
          /number=2
          /gene="rev"
          6202..8766
          /gene="env"
          /codon_start=6202
          8368..8449
          /number=3
          /gene="tat"
          8368..8750
          /number=3
          /gene="rev"
          8600..9271
          /gene="nef"
          /codon_start=8600
          8937..9623
          /note="3' long terminal repeat"
          9407..9623
          /rpt_type=R repeat 3' copy
          9506..9511
          /partial
BASE COUNT 3243 a 1807 c 2470 g 2103 t
ORIGIN    5' terminus of 5' LTR.

```



Query Match 57.9% Score 11; DB 3; Length 9623;  
Best Local Similarity 100.0%; Pred. No. 1.33e-01;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1513 tcattatgaa 1523  
|||||  
3 TCATTATGAA 13

RESULT 12  
LOCUS HIV260415K 781 bp ds-DNA VRL 23-AUG-1994  
DEFINITION Human immunodeficiency virus type 2 (HIV-2) p16/p28 gag sequence,  
sample 60415K, clone 8, partial cds.  
ACCESSION L33076  
SOURCE Human immunodeficiency virus type 2 PCR-amplified DNA recovered from  
a 34 year old heterosexual asymptomatic man from urban Senegal.  
ORGANISM Human immunodeficiency virus type 2  
REFERENCE Virdae; ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
1 (bases 1 to 781)  
AUTHORS Gao,F., Yue,L., Robertson,D.L., Hill,S.C., Hul,H., Biggar,R.J.,  
Hamn,B.H., Neeguaye,A.E., Whelan,T.M., Ho,D.D., Shaw,G.M., Sharp,P.M. and  
Hann,B.H.

TITLE Genetic diversity of human immunodeficiency virus type 2: Evidence  
for five distinct sequence subtypes with differences in virus  
biology  
J. Virol. 68, 7433-7447 (1994)

JOURNAL full staff-review  
STANDARD This sequence was obtained as part of a study designed to determine  
COMMENT the in vivo extent of HIV-2 genetic and biological variation. Up  
to now, HIV-2 sequences have been derived almost solely from  
cultured isolates. Since cultivation may select against viruses  
which can not grow in a particular cell type, Gao et al. argue that  
a greater variation may be elucidated by sequencing virus obtained  
from uncultured PBMCs. With this aim in mind, peripheral blood  
samples were collected from 12 HIV-2 seropositive patients (2238,  
60415K, 60667K, 7312A, 7810A, 7924A, FA, FO784, FT, JA, ON, and PA)  
from six different West African countries, living in both urban and  
rural areas. Uncultured PBMC DNA was PCR amplified, and the  
recombinant clones were manually sequenced. Another portion of the  
cells was used for virus isolation.  
Patient 60415K was a heterosexual 34-year old asymptomatic man from  
urban Senegal, who was living in the Washington, D.C. area. He  
traveled extensively before coming to the United States where he was  
identified as HIV-2 positive after seeking medical care for a STD.  
He was married to an American IV drug user. Virus isolation was  
successfully performed at the University of Alabama at Birmingham.  
Virus 60415K clusters with the subtype A HIV-2 viruses.  
NCBI gi: 532131

FEATURES  
source Location/Qualifiers  
1..781  
/organism="Human immunodeficiency virus type 2"  
/proviral  
/sequenced\_mol="DNA"  
/tissue\_type="PBMC"  
<1..>781  
/partial  
/gene="gag"  
/note="NCBI gi: 532132"  
/codon\_start=2  
/translation="ANELDFRGLAESLSEKSGCORLLTVLDPVTGSESLKSLFNT  
VCVAMCIIAEKVKDEGAKOIVORHLVAETGAEKSTSRPAPSGEGNPFVH  
VGVNVTIPLSPRLNMAVKIKERKGAJYVFOALSECTRPDINOMNCYGDHO  
AAMQIIRIETIEEAEMDVORPIPCPIPAQGLRPRSDIAGTISTYDEQIQMFRFO  
NPVPVGNIRWIGIGLQCKVRMNPINLIDIKGPREPFS"  
BASE COUNT 277 a 165 c 194 g 145 t  
ORIGIN 111 bp downstream from the start of gag

Query Match 52.6% Score 10; DB 3; Length 781;  
Best Local Similarity 100.0%; Pred. No. 1.82e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 509 attatgaag 518  
|||||  
10 ATTAATGAG 1

RESULT 13  
LOCUS SIVGR13E 1113 bp ss-RNA VRL 25-MAY-1994  
DEFINITION Simian immunodeficiency virus (SIVagm) partial envelope sequence,  
African green monkey isolate SIVagmRI-3.  
ACCESSION U03994  
SOURCE Simian immunodeficiency virus PCR-amplified DNA extracted from  
cocultured PBMCs derived from a grivet African green monkey from  
Ethiopia.  
ORGANISM Simian immunodeficiency virus  
REFERENCE Virdae; ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
1 (bases 1 to 1113)  
AUTHORS Jin,M.J., Hul,H., Robertson,D.L., Mueller,M.C., Barre-Sinoussi,F.,  
Hirsch,V.M., Allan,J.S., Shaw,G.M., Sharp,P.M. and Hann,B.H.  
Mosaic Genome Structure of Simian Immunodeficiency Virus from West  
African Green Monkeys  
EMBO J. 13, 2935-2947 (1994)

TITLE full staff-review  
JOURNAL This sequence was obtained as part of a study designed to determine  
STANDARD the breadth of genetic variation of SIV infecting African green  
COMMENT monkeys. With this aim in mind, Jin et al. PCR-amplified, cloned  
and sequenced the LTR and partial envelope region from 12 isolates  
(ver-1, ver-2, gri-2, gri-3, sab-1, sab-2, sab-3, sab-4, tan-1,  
tan-17, tan-40, and tan-43). These isolates were derived from  
different African green monkey species inhabiting various  
geographical regions. Most of the veret and grivet viruses and  
all of the sabous and tantalus viruses were derived from feral  
animals.  
NCBI gi: 466207

FEATURES  
source Location/Qualifiers  
1..1113  
/clone="envgri-3"  
/strain="SIVagmRI-3"  
/organism="simian immunodeficiency virus"  
/specific\_host="African green monkey"  
/note="grivet subtype"  
<1..>1113  
/note="NCBI gi: 466208"  
/codon\_start=1  
/product="envelope protein"  
/translation="TRAPRCACAGYALLKNDKDYNGFAPKCKNVSVCTLIINT  
ITTGICGNSRVNRETRIEWQKNSNDTVIANKRINLVRRRPNKTIITVTIHA  
GRVFSQKYNTRKLOAKCHSOGWMLGAKVREKIKELNLTVEKLENIHLRQWGP  
EANFVWNCHEFFCYCKMPFIYLNQITADAGKCTSHKSGGPGPCAKTYVACH  
IROVNDWMTLTKRYKAPREGHLECSNITVADALVEINNNKSGPIVTLTSQVSIM  
AYELGDVKEIETPIGAPFDVDRYRSPSEKRVPLGFLGLGAAGTGMGAATAVL  
TVOSRHLLAIILOOKNLVAVRQOOLKLTITWVK"

BASE COUNT 395 a 188 c 271 g 259 t  
ORIGIN approximately 727 bp downstream from beginning of env cds

Query Match 52.6% Score 10; DB 3; Length 1113;  
Best Local Similarity 100.0%; Pred. No. 1.82e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 540 gtcaataat 549  
|||||  
14 GTTCATTAT 5

RESULT 14  
LOCUS SIVAGMT17 1656 bp ds-DNA VRL 09-JUN-1993  
DEFINITION Simian immunodeficiency virus, isolate TAN-17, complete gag cds.  
ACCESSION L19250  
SOURCE Simian immunodeficiency virus proviral DNA, isolate TAN-9 from a  
Ugandan tantalus monkey (Ceopithecus tantalus).  
ORGANISM Simian immunodeficiency virus  
REFERENCE Virdae; ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.



REFERENCE 1 (bases 1 to 1696)  
AUTHORS Hirsch,V.M., McEann,C., Dapolito,G., Goldstein,S., Ogen-Odoi,A.,

TITLE Biryawaho,B., Lakwo,T. and Johnson,P.R.

JOURNAL Identification of a new subgroup of SIVagm in tantalus monkeys

STANDARD Virology (1993) In press

COMMENT full staff\_review

Kindly submitted prior to publication by Dr. Vanessa Hirsch,  
Immunodeficiency Viruses Section, NIAID/NIH, Twinbrook II,  
Rockville MD 20852. SIVagm-tan strains cluster with other  
previously characterized agm viral strains taken from vervets  
and grivets. See related tantalus viral isolate sequences 27,  
40, 49, and 9 (accession numbers L19251-L19254).

Location/Qualifiers

exon 143..1696

source /gene="gag"

/organism="Simian immunodeficiency virus"

/cell\_line="CEMss"

/haplotype="na"

/proviral

/sequenced\_mol="DNA"

BASE COUNT 581 a 362 c 432 g 321 t

ORIGIN 142 nt upstream from the gag cds start.

Query Match 52.6%; Score 10; DB 3; Length 1696;

Best Local Similarity 100.0%; Pred. No. 1.82e+00;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 779 ataatgaag 788

Cp 10 ATTATGAG 1

LOCUS 15 SIMSP2 3534 bp ss-RNA VRL 09-NOV-1990

DEFINITION Simian spumavirus, partial pol cds, complete env cds.

ACCESSION M33561

SEGMENT 2 of 2

SOURCE Simian spumavirus.

REFERENCE 1 (bases 1 to 3534)

AUTHORS Mergia,A., Shaw,K.E.S., Lackner,J.E. and Luciw,P.A.

TITLE Relationship of the env genes and the endonuclease domain of the es

of simian foamy virus type 1 and human foamy virus

J. Virol. 64, 406-410 (1990)

JOURNAL

STANDARD

COMMENT Kindly submitted in computer readable form by K.Shaw and P.Luciw

LTR start at base 4857 EcoRI to EcoRI fragment, lower case shows

area that was deleted in the genomic clone contains end of pol,

env.

FEATURES Location/Qualifiers

CDS <1..582

/note="polymerase (AA at 1)"

/codon\_start=1

524..3481

/note="envelope protein"

/codon\_start=524

BASE COUNT 1165 a 648 c 674 g 1047 t

ORIGIN

Query Match 52.6%; Score 10; DB 3; Length 3534;

Best Local Similarity 100.0%; Pred. No. 1.82e+00;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

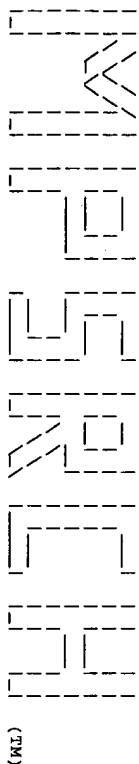
Db 1097 ataatgaag 1106

Cp 10 ATTATGAG 1

Search completed: Sat Nov 27 15:14:40 1999

Job time : 6 secs.





(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPearch\_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 15:13:27 1999; MasPar time 19.75 Seconds

Tabular output not generated. 206.098 Million cell updates/sec

Title: >US-09-103-287-5  
Description: (1-19) from US09103287.seq  
Perfect Score: 19  
N.A. Sequence: 1 CTTCATTATGACGATGC 19  
Comp: GAAGTATATCTGCTACG

Scoring table: TABLE jmetric  
Gap 60

Mmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

n-geneseq35  
1:part1.2:part2.3:part3.4:part4.5:part5.6:part6.7:part7.8:part8.9:part9.10:part10.11:part11.12:part12.13:part13.14:part14.15:part15.16:part16.17:part17.18:part18.19:part19.20:part20.21:part21.22:part22.23:part23.24:part24.25:part25.26:part26.27:part27.28:part28.29:part29.30:part30.31:part31.32:part32.33:part33.34:part34.35:part35.36:part36.37:part37.38:part38.39:part39.40:part40.41:part41.42:part42.43:part43.44:part44.45:part45.46:part46.47:part47.48:part48.49:part49.50:part50.51:part51.52:part52.53:part53.54:part54.55:part55.56:part56.57:part57.58:part58.59:part59.60:part60

Statistics: Mean 4.607; Variance 2.283; scale 2.018

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	19	100.0	19	60	Murc polynucleotides	2.76e-03
2	19	100.0	619	V53479	DNA encoding a Staphy	2.76e-03
3	19	100.0	660	V80065	Partial nucleotide se	2.76e-03
4	19	100.0	1351	V99650	UDP-N-acetylmuramate: 5.81e+00	2.76e-03
5	14	73.7	517	43	DNA encoding a Staphy	2.43e+01
6	13	68.4	743	2	Sequence encoding the	2.43e+01
7	13	68.4	1113	35	Mangifera indica ACC	2.43e+01
8	13	68.4	1138	9	Human GM-CSF gene.	2.43e+01
9	13	68.4	1707	3	Secretion enhancing s	2.43e+01

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
10	13	68.4	3043	43	V29700	Human GM-CSF encoding
11	13	68.4	3194	10	Q58039	GM-CSF gene.
12	13	68.4	4846	58	V74587	Staphylococcus aureus
13	13	68.4	24004	60	X13009	Enterococcus faecalis
14	13	68.4	235033	51	V57926	Hereditary haemochrom
15	13	68.4	237326	50	V57903	Hereditary haemochrom
16	12	63.2	19	60	V99651	Murc polynucleotides
17	12	63.2	211	57	V77471	Staphylococcus aureus
18	12	63.2	287	54	V71830	Subgenomic fragment o
19	12	63.2	306	2	N70293	Sequence encoding the
20	12	63.2	400	57	V76242	Staphylococcus aureus
21	12	63.2	400	8	O60125	Human brain Expressed
22	12	63.2	594	1	Q06816	Bovine Leukaemia viru
23	12	63.2	619	48	V53479	DNA encoding a Staphy
24	12	63.2	730	33	T85435	Human cadherin-10 cod
25	12	63.2	730	28	T61924	Partial human cadheri
26	12	63.2	730	33	T85435	Human cadherin-10 cod
27	12	63.2	730	9	O51325	Human cadherin-10 gen
28	12	63.2	730	28	T61924	Partial human cadheri
29	12	63.2	743	2	N70260	Sequence encoding the
30	12	63.2	1098	58	V74564	Staphylococcus aureus
31	12	63.2	1138	9	O55627	Human GM-CSF gene.
32	12	63.2	1268	60	X13494	Enterococcus faecalis
33	12	63.2	1521	40	V10132	Human GM-CSF cDNA.
34	12	63.2	1789	23	T34657	NIA protease of the F
35	12	63.2	3775	58	V74549	Staphylococcus aureus
36	12	63.2	6536	40	V04471	Nudarella beta-like
37	12	63.2	7587	40	V15703	Zucchini ACC synthase
38	12	63.2	7587	3	O15133	Zucchini ACC synthase
39	12	63.2	7616	12	O70194	Vector pOTS208NS1Bimu
40	12	63.2	7646	58	V74432	Staphylococcus aureus
41	12	63.2	10690	47	V52226	Streptococcus pneumon
42	12	63.2	10690	47	V52226	Streptococcus pneumon
43	12	63.2	13585	17	T11549	Tumour rejection anti
44	12	63.2	13747	8	O47342	Myotonic dystrophy ge
45	12	63.2	237326	50	V57903	Hereditary haemochrom

## ALIGNMENTS

RESULT 1  
ID V99651 standard; DNA; 19 BP.  
AC V99651;  
DT 17-MAR-1999 (first entry)  
DE Murc polynucleotides amplifying primer.  
KW Murc gene; UDP-N-acetylmuramate:L-alanine ligase; Murc polypeptide;  
KW bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
KW immunogen; drug; genetic immunisation; PCR primer; ss.  
OS Synthetic.  
OS Staphylococcus aureus.  
PN EP-889123-A2.  
PD 07-JAN-1999.  
PF 26-JUN-1998; 305064.  
PR 03-JUL-1997; US-052720.  
PA (SMK ) SMITHKLINE BEECHAM CORP.  
PA (SMK ) SMITHKLINE BEECHAM PLC.  
PI Burnham MKR, Wallis NG;  
DR WPI; 99-062655/06.  
PT New isolated Murc polypeptide from Staphylococcus aureus and related  
PT nucleic acid - useful in diagnosis, treatment and prevention of  
PT bacterial infections  
PS Disclosure; Page 13; 39pp; English.  
CC The invention relates to a UDP-N-acetylmuramate:L-alanine ligase (Murc  
CC polypeptide) encoded by the S. aureus Murc gene. Host cells containing  
CC an expression system comprising the Murc gene can be used for the  
CC recombinant production of the polypeptide. Agonists or the Murc  
CC polypeptide are used to treat conditions requiring increased activity or  
CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or  
CC competitive polypeptide are useful for inhibiting the polypeptide e.g.  
CC bacterial (especially S. aureus) infections. They are also useful against  
CC Helicobacter pylori infections and related cancers, ulcers and gastritis.  
CC The antibacterial agents are useful to treat in-dwelling devices for  
CC infection prevention or generally as wound treatments to prevent adhesion



CC of bacteria to matrix proteins. The MurC polypeptide is also useful for  
 CC diagnosing or prognosing a (susceptibility to) disease, for raising  
 CC antibodies, to identify modulators or specific receptors; in rational  
 CC drug design and as an immunogen for vaccines. The MurC gene sequences are  
 CC useful in antisense/ribozyme therapeutics; to detect mutant MurC gene;  
 CC for chromosomal mapping; to determine bacterial serotype; and for genetic  
 CC immunisation. The present sequence represents a primer used for the PCR  
 CC amplification of the MurC polynucleotides.  
 CC Sequence 19 BP: 6 A; 4 C; 3 G; 6 T;

Query Match 100.0%; Score 19; DB 60; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 2.76e-03;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 ctccatgaacgagtc 19  
 |||||  
 1 CTTCATTATGAACGATGC 19

RESULT 2  
 ID V53479 standard; DNA: 619 BP.

AC V53479;  
 DT 30-OCT-1998 (first entry)  
 DE DNA encoding a Staphylococcus aureus protein of unknown function.  
 KW Staphylococcus aureus protein; immune response induction; eye infection;  
 KW antibody production; T-cell immune response; gastrointestinal infection;  
 KW respiratory infection; inhibitor; bacterial infection; cardiac infection;  
 KW central nervous system; kidney infection; urinary tract infection;  
 KW antimicrobial compound identification; broad spectrum antibiotic;  
 KW therapy; ss.  
 OS Staphylococcus aureus.  
 PN EP-841394-A2.  
 PD 13-MAY-1998.  
 PE 24-SEP-1997; 307485.  
 PR 24-SEP-1996; US-027032.

PA (SMK ) SMITHKLINE BEECHAM CORP.  
 (SMK ) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Burnham MKR, Hodgson JE, Knowles DJC,  
 PI Lometto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,  
 PI Ward JM;  
 PI WPI: 98-252940/23.  
 DR P-PSDB: W77686.

PT New nucleic acid sequences from Staphylococcus aureus WCHU29 -  
 PT useful in vaccines and for treatment of bacterial infections of e.g.  
 PT respiratory tract and central nervous system  
 PS Claim 1; Page 146; 390pp: English.

CC This sequence encodes a Staphylococcus aureus protein of unknown  
 CC function, and represents a DNA Sequence of the invention.  
 CC The DNA sequences were isolated from Staphylococcus aureus WCHU29  
 CC (NCIMB 40771). Host cells containing the DNA sequences are used to  
 CC produce polypeptides or fragments. The proteins are used in the treatment  
 CC of disease, for inducing an immune response by administering them, to  
 CC produce antibody and/or T-cell immune response. Antagonists of the  
 CC proteins are used for the inhibition of bacterial polypeptides.  
 CC Conditions which may be treated include bacterial infections, especially  
 CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,  
 CC urinary tract, skin, bones and joints. The proteins can also be used to  
 CC identify anticancer compounds which are broad spectrum antibiotics,  
 CC especially useful in the treatment of H. pylori infection.  
 CC Sequence 619 BP: 208 A; 117 C; 79 G; 215 T;

Query Match 100.0%; Score 19; DB 48; Length 619;  
 Best Local Similarity 100.0%; Pred. No. 2.76e-03;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 137 ctccatgaacgagtc 155  
 |||||  
 1 CTTCATTATGAACGATGC 19

RESULT 3  
 ID V80065 standard; DNA: 660 BP.

AC V80065;  
 DT 17-MAR-1999 (first entry)  
 DE Partial nucleotide sequence of the MurC gene.  
 KW MurC gene; UDP-N-acetylmuramate:L-alanine ligase; MurC polypeptide;  
 KW bacterial infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
 KW immunogen; drug; genetic immunisation; ds.  
 OS Staphylococcus aureus.  
 FH Key  
 FT Location/Qualifiers  
 FT CDS  
 FT 2..660  
 FT /tag= a  
 FT /product= "partial MurC polypeptide"

PN EP-889123-A2.  
 PD 07-JAN-1999.  
 PE 26-JUN-1998; 305064.  
 PR 03-JUL-1997; US-052720.  
 PA (SMK ) SMITHKLINE BEECHAM CORP.  
 PA (SMK ) SMITHKLINE BEECHAM PLC.  
 PI Burnham MKR, Wallis NG;  
 PI WPI: 99-06265/06.  
 DR P-PSDB: W89199.

PT New isolated MurC polypeptide from Staphylococcus aureus and related  
 PT nucleic acid - useful in diagnosis, treatment and prevention of  
 PT bacterial infections  
 PS Claim 2; Pages 4-5; 39pp; English.

CC The invention relates to a UDP-N-acetylmuramate:L-alanine ligase  
 CC (MurC polypeptide) encoded by the S. aureus MurC gene. Host cells  
 CC containing an expression system comprising the MurC gene can be used for  
 CC the recombinant production of the polypeptide. Agonists or the MurC  
 CC polypeptide are used to treat conditions requiring increased activity or  
 CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or  
 CC competitive polypeptide are useful for inhibiting the polypeptide e.g.  
 CC bacterial (especially S. aureus) infections. They are also useful against  
 CC Helicobacter pylori infections and related cancers, ulcers and gastritis.  
 CC The antibacterial agents are useful to treat in-dwelling devices for  
 CC infection prevention or generally as wound treatments to prevent adhesion  
 CC of bacteria to matrix proteins. The MurC polypeptide is also useful for  
 CC diagnosing or prognosing a (susceptibility to) disease, for raising  
 CC antibodies, to identify modulators or specific receptors; in rational  
 CC drug design and as an immunogen for vaccines. The MurC gene sequences are  
 CC useful in antisense/ribozyme therapeutics; to detect mutant MurC gene;  
 CC for chromosomal mapping; to determine bacterial serotype; and for genetic  
 CC immunisation. The present sequence represents a partial nucleotide  
 CC sequence (MurC ORF) of the MurC gene.  
 CC Sequence 660 BP: 233 A; 83 C; 120 G; 224 T;

Query Match 100.0%; Score 19; DB 60; Length 660;  
 Best Local Similarity 100.0%; Pred. No. 2.76e-03;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 512 gcatcgcttcattatgaag 530  
 |||||  
 19 GCATCGCTTCAATTATGAAG 1

RESULT 4  
 ID V99650 standard; DNA: 1351 BP.

AC V99650;  
 DT 17-MAR-1999 (first entry)  
 DE UDP-N-acetylmuramate:L-alanine ligase (MurC polypeptide) encoding DNA.  
 KW MurC gene; UDP-N-acetylmuramate:L-alanine ligase; MurC polypeptide;  
 KW bacterial infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
 KW immunogen; drug; genetic immunisation; ds.  
 OS Staphylococcus aureus.

FH Key  
 FT Location/Qualifiers  
 FT CDS  
 FT 22..1335

FT /tag= a  
 FT /gene= "MurC"  
 FT /product= "UDP-N-acetylmuramate:L-alanine ligase  
 (MurC polypeptide)"

PN EP-889123-A2.  
 PD 07-JAN-1999.  
 PE 26-JUN-1998; 305064.  
 PR 03-JUL-1997; US-052720.



PA (SMIR ) SMITHKLINE BEECHAM CORP.  
 PA (SMIR ) SMITHKLINE BEECHAM PLC.  
 PI Burnham MKR, Wallis NG;  
 DR WPI: 99-062655/06.  
 DR P-PSDB: W87771.  
 PT New isolated MucC polypeptide from *Staphylococcus aureus* and related  
 PT nucleic acid - useful in diagnosis, treatment and prevention of  
 PT bacterial infections  
 PS Claim 2: Pages 3-4: 39pp; English.  
 CC The present sequence represents a MucC gene encoding a *Staphylococcus*  
 CC aureus UDP-N-acetylmuramyl-L-alanine ligase (MucC polypeptide). Host  
 CC cell containing an expression system comprising the MucC gene can be used  
 CC for the recombinant production of the polypeptide. Agonists or the MucC  
 CC polypeptide are used to treat conditions requiring increased activity or  
 CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or  
 CC competitive polypeptide are useful for inhibiting the polypeptide e.g.  
 CC bacterial (especially *S. aureus*) infections. They are also useful against  
 CC *Helicobacter pylori* infections and related cancers, ulcers and gastritis.  
 CC The antibacterial agents are useful to treat in-dwelling devices for  
 CC infection prevention or generally as wound treatments to prevent adhesion  
 CC of bacteria to matrix proteins. The MucC polypeptide is also useful for  
 CC diagnosing or prognosing a (susceptibility to) disease, for raising  
 CC antibodies to identify modulators or specific receptors; in rational  
 CC drug design and as an immunogen for vaccines. The MucC gene sequences are  
 CC useful in antisense/ribozyme therapeutics; to detect mutant MucC gene;  
 CC for chromosomal mapping; to determine bacterial serotype; and for genetic  
 CC immunisation.  
 SQ Sequence 1351 BP; 471 A; 166 C; 260 G; 454 T;

Query Match 100.0%; Score 19; DB 60; Length 1351;  
 Best Local Similarity 100.0%; Pred. No. 2.76e-03;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1198 gcatcgtcattatgaag 1216  
 |||  
 CP 19 GCATCGTCATTATGAAG 1

RESULT 5  
 ID T83852 standard; DNA; 517 BP.  
 AC T83852;  
 DT 20-AUG-1998 (first entry)  
 DE DNA encoding a *Staphylococcus aureus* protein of unknown function.  
 KW *Staphylococcus aureus* protein; ribozyme; antisense sequence; control;  
 KW *Staphylococcus aureus* gene; regulatory element; bacterial gene expression;  
 KW vaccine; *Staphylococcus aureus* infection; food poisoning; scaled skin syndrome;  
 KW toxic shock syndrome; ss.  
 OS *Staphylococcus aureus*.  
 FH Location/Qualifiers  
 FT Key  
 FT CDS Complement (296..424)  
 FT /\*tag- a

MO9730070-A1.  
 PD 21-AUG-1997.  
 PF 19-FEB-1997; 002318.  
 PR 20-FEB-1996; US-011888.  
 PA (SMIR ) SMITHKLINE BEECHAM CORP.  
 PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,  
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;  
 DR WPI: 97-424969/39.  
 DR P-PSDB: W27890.  
 PT Novel polypeptide(s) from *Staphylococcus aureus* strain WC029 - used  
 PT to isolate antimicrobial compounds, and in vaccines against *S.*  
 PT aureus infection  
 PS Claim 9: Page 711: 989pp; English.  
 CC The present sequence encodes a *Staphylococcus aureus* protein of  
 CC unknown function. The present sequence was isolated from a  
 CC library of clones of *S. aureus* WC029 in *Escherichia coli*. The DNA  
 CC sequence can be used in the construction of ribozymes and antisense  
 CC sequences to control the expression of *Staphylococcus* genes. The DNA  
 CC sequence is also useful as a source of regulatory elements for the  
 CC control of bacterial gene expression. The encoded protein may be used  
 CC to produce vaccines to enable a host to produce specific antibodies  
 CC with antibacterial action. These vaccines and antibodies would protect

CC a host against invasion by *S. aureus*, and conditions relating to  
 CC *Staphylococcus aureus* infection, e.g. *Staphylococcus aureus* food poisoning, scaled  
 CC skin syndrome, and toxic shock syndrome.  
 SQ Sequence 517 BP; 131 A; 101 C; 75 G; 194 T;

Query Match 73.7%; Score 14; DB 43; Length 517;  
 Best Local Similarity 93.3%; Pred. No. 5.81e+00;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 26 atccttataatga 40  
 |||  
 CP 17 ATCCTTATATGA 3

RESULT 6  
 ID N70260 standard; DNA; 743 BP.  
 AC N70260;  
 DT 03-APR-1991 (first entry)  
 DE Sequence encoding the 46 amino acid product of the *sacQ* gene of *B.*  
 DE *amyloidifaciens* called 'enhancing DNA sequence (EDS) Apro'.  
 KW Gram-positive bacteria; gene expression; ss.  
 OS *Bacillus amyloidifaciens*.  
 EP-227260-A1.  
 PD 01-JUL-1987.  
 PF 27-OCT-1986; 308356.  
 PR 25-OCT-1985; US-791350.  
 PR 20-OCT-1986; US-921343.  
 PA (BIOT-) BIOTECHNICA INT. INC.  
 PI Sloma A, Lee RC, Pero J.  
 DR WPI: 87-179522/26.  
 PT Enhancing expression of genes in bacterial cells - by  
 PT transformation with enhancing DNA sequence which enhances prodn.  
 PS of endogenous protease  
 PS disclosure: Fig 2A: 39pp; English.  
 CC The EDS is pref. carried on a multicopy plasmid and/or is  
 CC integrated into the chromosome of the bacterial cell which is most  
 CC pref. of a *B. subtilis* strain. The cells provide enhanced  
 CC expression of desired genes in bacteria such as *B. subtilis* which  
 CC does not require modifications in the fermentation conditions of the  
 CC bacteria and which does not affect the properties of the desired end-  
 CC prodn.  
 SQ Sequence 743 BP; 261 A; 119 C; 138 G; 225 T;

Query Match 68.4%; Score 13; DB 2; Length 743;  
 Best Local Similarity 100.0%; Pred. No. 2.43e+01;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 414 ctccataatga 426  
 |||  
 QY 1 CTCCTATTATGA 13

RESULT 7  
 ID T72632 standard; DNA; 1113 BP.  
 AC T72632;  
 DT 27-JAN-1998 (first entry)  
 DE *Mangifera indica* ACC synthase m1acc2 gene.  
 KW ACC synthase; S-adenosyl-L-methionine methylthioadenosine lyase;  
 KW m1acc2; ethylene biosynthesis; transgenic plant; senescence;  
 KW antisense expression system; plant development; fruit ripening;  
 KW EC 4.4.1.14; mango; multigene family; ss.  
 OS *Mangifera indica*.  
 FH Key  
 FT Location/Qualifiers  
 FT CDS 1..1113  
 FT /\*tag- a  
 FT /EC\_number= 4.4.1.14  
 FT /product= m1acc2  
 FT /note= "Sequence represents 75% of the coding  
 FT sequence and does not contain the start  
 FT or stop codons."

MO9711166-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; AU0591.



PR 02-MAY-1996: AU-009603.  
PR 20-SEP-1995: AU-005559.  
PA (UYOU) UNIV QUEENSLAND.  
PI Botella JR.  
DR WPI: 97-202875/18.  
DR P-PSDB: W18290.  
PT Pineapple, papaya and mango ACC synthase genes - used in gene  
PT therapy to produce fruits with reduced senescence  
CC Claim 5, Fig 5, 46pp: English.  
CC This sequence represents the novel gene, macc2, which is a member  
CC of the mango ACC synthase multigene family. ACC synthase is an enzyme  
CC involved in the pathway for ethylene biosynthesis and the rate of  
CC endogenous expression of ACC synthase is considered to limit  
CC substantially the rate of ethylene production. Endogenous ethylene is  
CC often deleterious to crops, especially if some form of mechanical  
CC wounding has occurred and diminishes their post harvest quality and  
CC storage life. Novel ACC synthase genes expressed in transgenic plants  
CC using either sense or antisense expression system may be used to control  
CC the regulation of plant development, in particular fruit ripening,  
CC reducing senescence and thus improving storage life.  
SQ Sequence 1113 BP; 320 A; 212 C; 264 G; 317 T;

Query Match 68.4%; Score 13; DB 35; Length 1113;  
Best Local Similarity 100.0%; Pred. No. 2.43e+01;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 531 cttcataatgaa 543  
|||||  
QY 1 CTTCAATTAATGAA 13

RESULT 8  
ID Q55627 standard; DNA: 1138 BP.  
AC Q55627;  
DT 12-JUL-1994 (first entry)  
DE Human GM-CSF gene.  
KW Granulocyte Macrophage Colony Stimulating Factor; GM-CSF;  
KW non-coding region; coding region; resonance; interaction;  
OS optimisation; promoter region; TATA box; ss.  
PN Homo sapiens.  
FN FR2692594-A.  
PD 24-DEC-1993.  
PF 22-JUN-1992: 007571.  
PR 22-JUN-1992; FR-007571.  
PA (PERE/) PEREZ J.  
PI Perez J;  
DR WPI: 94-078256/04.  
PT Application of optimised gene expression - for scientific.  
PT Industrial and therapeutic purposes  
PS Disclosure: Fig 13; 110pp: French.  
CC Alterations were made to the TATA box sequence of the human and  
CC mouse GM-CSF genes (see Q55627 and Q55628, respectively) and the  
CC effects of the various alterations on gene expression were measured.  
CC Results indicated that resonances of the LFR type are most strongly  
CC affected by changes in the TATA box; it was further shown that  
CC LFR resonances control expression.  
SQ Sequence 1138 BP; 274 A; 306 C; 292 G; 266 T;

Query Match 68.4%; Score 13; DB 9; Length 1138;  
Best Local Similarity 100.0%; Pred. No. 2.43e+01;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 177 gttcataatgaa 189  
|||||  
Cp 14 GTTCATTAATGAA 2

RESULT 9  
ID N50411 standard; DNA: 1707 BP.  
AC N50411;  
DT 20-JAN-1992 (first entry)  
DE Secretion enhancing sequence.  
KW Protease; Bacillus; ss.

OS Bacillus amyloliquefaciens.  
PN J60058076-A.  
PD 04-APR-1985.  
PF 12-SEP-1983: 16665.  
PR 12-SEP-1983; JP-16665.  
PA (AGEN) AGENCY OF IND SCI TECH.  
PA (NITT) MIN INT TRADE & IND.  
DR WPI: 85-118649/20.  
DR DNA base sequence - with improved protein-secretory power.  
PS Claim 1; Page 425-426; 8pp: Japanese.  
CC The sequence enhances associated protein expression, and may thus  
CC act as a useful marker for recombinant plasmid expression.  
SQ Sequence 1707 BP; 505 A; 342 C; 398 G; 462 T;

Query Match 68.4%; Score 13; DB 3; Length 1707;  
Best Local Similarity 100.0%; Pred. No. 2.43e+01;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1436 cttcataatgaa.1448  
|||||  
QY 1 CTTCAATTAATGAA 13

RESULT 10  
ID V29700 standard; DNA: 3043 BP.  
AC V29700;  
DT 18-AUG-1998 (first entry)  
DE Human GM-CSF encoding DNA.  
KW Recombinant poxvirus; biological response modifier; IL-12; GM-CSF;  
KW granulocyte macrophage colony stimulating factor; prevention;  
KW tumour; interleukin; cytokine; treatment; ss.  
OS Homo sapiens.  
FN Location/Qualifiers  
FH Key  
FT TATA\_signal 597..603  
FT /\*tag= a  
FT 662..2685  
FT /\*tag= b  
FT /product= "Human GM-CSF"  
FT /note= "contains introns"  
FT 662..820  
FT /\*tag= c  
FT /number= 1  
FT 821..913  
FT /\*tag= d  
FT /number= 1  
FT 914..955  
FT /\*tag= e  
FT /number= 2  
FT 956..1642  
FT /\*tag= f  
FT /number= 2  
FT 1643..1768  
FT /\*tag= g  
FT /number= 3  
FT 1769..2577  
FT /\*tag= h  
FT /number= 3  
FT 2578..2682  
FT /\*tag= i  
FT /number= 4

FT W09815635-A2.  
FT 16-APR-1998.  
FT PD 03-OCT-1997; CA0729.  
FT PR 04-OCT-1996; US-027530.  
FT PA (CONN-) CONNAUGHT LAB LTD.  
FT (VITRO-) VIROGENETICS CORP.  
FT Alexander J, Bonnet MC, Cornet B, Cox B, Diaz I,  
FT Gajewczyk DM, Klein MH, Paoletti E, Tartaglia J;  
FT WPI: 98-240821/21.  
DR P-PSDB: M60031.  
DR Recombinant poxvirus encoding biological response modifier - useful  
PT for the treatment or prevention of tumours  
PS Disclosure; Fig 9A-F; 52pp: English.



CC This DNA encodes the human granulocyte macrophage colony stimulating factor (GM-CSF). A murine GM-CSF which is similar to the human GM-CSF is used for generating a recombinant poxvirus. The recombinant poxvirus comprises a nucleic acid sequence encoding a biological response modifier or its immunomodulating fragment in a non-essential region of the CC recombinant virus genome. The biological response modifier may be a CC cytokine, e.g. Interleukin-12 (IL-12), IL-2 or GM-CSF. The recombinant CC poxvirus is used to transfect cells for expression of the biological CC response modifier. The products are useful for the treatment of tumours CC or for protecting a host against tumour formation.

SO Sequence 3043 BP; 668 A; 824 C; 878 G; 673 T;

Query Match 68.4%; Score 13; DB 43; Length 3043;  
Best Local Similarity 100.0%; Pred. No. 2.43e+01;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 447 gtccataatgaa 459  
Cp 14 GTTCATTATGAA 2

RESULT 11  
ID 058039 standard; cDNA to mRNA; 3194 BP.  
AC 058039; (first entry)  
DT 01-AUG-1994  
DE GM-CSF gene.  
KW granulocyte-macrophage colony stimulating factor; GM-CSF; cytokine;  
KM tumor therapy; gene therapy; tumor-specific promoter; ds.  
OS Homo sapiens.  
PN M09404196.A.  
PD 03-MAR-1994.  
PF 16-AUG-1993; G01730.  
PR 14-AUG-1992; GB-017270.  
PR 27-FEB-1993; GB-004024.  
PA (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.  
PI Hart IR, Vile RG;  
DR WPI; 94-082848/10.  
PT DNA construct encoding cytokine for expression in tumour cells -  
PS for treating melanoma, pancreatic, breast, colon, prostate cancer  
PT disclosure; Page 80-82; 107pp; English.  
CC Cytokine genes/ e.g the GM-CSF gene, are attached to a  
CC tumor-specific promoter and delivered to patients for tumor gene  
CC therapy.  
CC Sequence 3194 BP; 701 A; 858 C; 945 G; 690 T;

Query Match 68.4%; Score 13; DB 10; Length 3194;  
Best Local Similarity 100.0%; Pred. No. 2.43e+01;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 450 gtccataatgaa 462  
Cp 14 GTTCATTATGAA 2

RESULT 12  
ID V74587 standard; DNA; 4846 BP.  
AC V74587;  
DT 16-MAR-1999 (first entry)  
DE Staphylococcus aureus contig SEQ ID #276.  
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KM skin infection; surgical wound infection; scalded skin syndrome;  
KM toxic shock syndrome; ds.  
OS Staphylococcus aureus  
FH key Location/Qualifiers  
FT misc-feature 1261..1320  
FT /tag= a  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

PN EP-786519-A2.  
PD 30-JUL-1997.  
PF 07-JAN-1997; 100117.  
PR 05-JAN-1996; US-009861.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
PI Rosen CA;  
DR WPI; 97-374922/35.  
FT Polynucleotide(s) and proteins derived from Staphylococcus aureus  
PT stored on computer readable medium and used in the production of  
PT anti-S.aureus vaccines  
PS Claim 1; Page 1092-1094; 327pp; English.  
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
CC of the invention. The DNA sequences are recorded on a computer readable  
CC medium, preferably selected from a floppy or hard disk, random access  
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
CC the S.aureus DNA sequences allows putative functions to be assigned so  
CC that protein-encoding or regulatory regions of commercial, therapeutic or  
CC industrial importance can be obtained. Specifically, sequences which are  
CC likely to encode antigens have been identified and these polypeptides can  
CC be used in a vaccine composition against S.aureus infection. The  
CC polypeptides can also be used in a kit for the immunodetection of  
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
CC for recombinant production of the polypeptides. The new DNA sequences  
CC (and their fragments) are useful as primers or probes for isolating  
CC homologues of any of the S.aureus DNA sequences contained on the  
CC computer readable medium.

SO Sequence 4846 BP; 1551 A; 869 C; 650 G; 1656 T;

Query Match 68.4%; Score 13; DB 58; Length 4846;  
Best Local Similarity 100.0%; Pred. No. 2.43e+01;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 597 ttccataatgaa 609  
Cp 13 TTCCATTATGAA 1

RESULT 13  
ID X13009 standard; DNA; 24004 BP.  
AC X13009;  
DT 19-MAR-1999 (first entry)  
DE Enterococcus faecalis genome contig SEQ ID NO:72.  
KW Enterococcus faecalis; contig; detection; Enterococcal infection;  
KM vaccine; attenuation; computer readable medium; ds.  
OS Enterococcus faecalis.  
PN M09850555-A2.  
PD 12-NOV-1998.  
PF 04-MAY-1998; U08985.  
PR 14-NOV-1997; US-006609.  
PR 06-MAY-1997; US-044031.  
PR 16-MAY-1997; US-046653.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Dillon PJ, Kunsch CA;  
DR WPI; 99-045171/04.  
PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
PT infection.  
PS Claim 1; Page 526-538; 2084pp; English.  
CC A computer readable medium has been developed which has recorded on it  
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC X12938 to X13919 represent these nucleotide sequences which are primary  
CC nucleotide sequences, also known as contigs. The computer-based system  
CC can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence



CC of Enterococcus faecalis in samples. They can also be used for  
 CC diagnosing Enterococcal infection in an animal and monitoring  
 CC progression of disease, and for identifying agents which can be used to  
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
 CC another related organism, in vivo or in vitro. In particular the  
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
 CC can be used in vaccines to prevent or attenuate an Enterococcal  
 CC infection.  
 SQ Sequence 24004 BP; 8181 A; 3870 C; 4564 G; 7291 T;

Query Match 68.4%; Score 13; DB 60; Length 24004;  
 Best Local Similarity 100.0%; Pred. No. 2.43e+01;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 9057 ttcataatgaag 9069  
 ||||||||||||  
 QY 3 TCATTATGACG 15

RESULT 14  
 ID V57926 standard; DNA; 235033 BP.  
 AC V57926;  
 DT 23-DEC-1998 (first entry)  
 DE Hereditary haemochromatosis subregion from an unaffected individual.  
 KW Bovine butyrophilin; BT: human hereditary haemochromatosis; HFE;  
 CC diagnosis; iron metabolism; NPT3; NPT4; Roret; BTF1; BTF2; BTF3;  
 KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;  
 KM type 1 sodium transport gene; ss.  
 OS Homo sapiens.  
 PN MO9814466-A1.  
 PD 09-APR-1998.  
 PF 30-SEP-1997; U17658.  
 PR 07-MAY-1997; US-852495.  
 PR 01-OCT-1996; US-724394.  
 PA (PROC-) PROAGENTOR INC.  
 PI Feder JN, Krommal GS, Lauer PM, Ruddy DA, Thomas WJ,  
 PI Tsuchihashi Z, Wolff RK;  
 DR WPI; 98-240014/21.  
 PT Hereditary haemochromatosis gene products - used to develop products  
 PT for the diagnosis and treatment of hereditary disorders in iron  
 PT metabolism  
 PS Example 2; Fig 8; 209pp; English.  
 CC The present invention describes hereditary haemochromatosis gene  
 CC products from the human haemochromatosis gene. The present sequence  
 CC represents a hereditary haemochromatosis subregion from an individual  
 CC unaffected by hereditary haemochromatosis (HH). Also described is a  
 CC method to determine the presence or absence of the common hereditary  
 CC haemochromatosis (HFE) gene mutation in an individual comprising:  
 CC (a) providing DNA or RNA from the individual; and (b) assessing the  
 CC DNA or RNA for the presence or absence of a haplotype or genotype where  
 CC the presence or absence of the haplotype genotype indicates the likely  
 CC presence of the HFE gene mutation in the genome of the individual. The  
 CC HFE gene sequences from the present invention can be used to develop  
 CC products for use in the diagnosis and treatment of HFE. The present  
 CC invention also describes BTF genes, which are homologues of the milk  
 CC protein butyrophilin (BT), and can be used in the production of agonists  
 CC and antagonists of BT function. Also described are: (1) a Roret gene  
 CC which can be used to develop products for the study, diagnosis and  
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes  
 CC which are homologues of a type 1 sodium transport gene, and can  
 CC similarly be used for hypophosphatemia.  
 SQ Sequence 235033 BP; 68786 A; 48466 C; 49441 G; 68340 T;

Query Match 68.4%; Score 13; DB 51; Length 235033;  
 Best Local Similarity 100.0%; Pred. No. 2.43e+01;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 74200 ttcataatgaag 74212  
 ||||||||||||  
 Cp 13 TTCATTATGACG 1

RESULT 15

ID V57903 standard; DNA; 237326 BP.

AC V57903;  
 DT 21-DEC-1998 (first entry)  
 DE Hereditary haemochromatosis subregion from an HH affected individual.  
 KW Bovine butyrophilin; BT: human hereditary haemochromatosis; HFE;  
 CC diagnosis; iron metabolism; NPT3; NPT4; Roret; BTF1; BTF2; BTF3;  
 KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;  
 KM type 1 sodium transport gene; ss.  
 OS Homo sapiens.  
 PN MO9814466-A1.  
 PD 09-APR-1998.  
 PF 30-SEP-1997; U17658.  
 PR 07-MAY-1997; US-852495.  
 PR 01-OCT-1996; US-724394.  
 PA (PROC-) PROAGENTOR INC.  
 PI Feder JN, Krommal GS, Lauer PM, Ruddy DA, Thomas WJ,  
 PI Tsuchihashi Z, Wolff RK;  
 DR WPI; 98-240014/21.  
 PT Hereditary haemochromatosis gene products - used to develop products  
 PT for the diagnosis and treatment of hereditary disorders in iron  
 PT metabolism  
 PS Claim 1; Fig 9; 209pp; English.

CC The present invention describes hereditary haemochromatosis gene  
 CC products from the human haemochromatosis gene. The present sequence  
 CC represents a hereditary haemochromatosis subregion from an hereditary  
 CC haemochromatosis (HH) affected individual. Also described is a  
 CC method to determine the presence or absence of the common hereditary  
 CC haemochromatosis (HFE) gene mutation in an individual comprising:  
 CC (a) providing DNA or RNA from the individual; and (b) assessing the  
 CC DNA or RNA for the presence or absence of a haplotype or genotype where  
 CC the presence or absence of the haplotype genotype indicates the likely  
 CC presence of the HFE gene mutation in the genome of the individual. The  
 CC HFE gene sequences from the present invention can be used to develop  
 CC products for use in the diagnosis and treatment of HFE. The present  
 CC invention also describes BTF genes, which are homologues of the milk  
 CC protein butyrophilin (BT), and can be used in the production of agonists  
 CC and antagonists of BT function. Also described are: (1) a Roret gene  
 CC which can be used to develop products for the study, diagnosis and  
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes  
 CC which are homologues of a type 1 sodium transport gene, and can  
 CC similarly be used for hypophosphatemia.  
 SQ Sequence 237326 BP; 69596 A; 48904 C; 48217 G; 70609 T;

Query Match 68.4%; Score 13; DB 50; Length 237326;  
 Best Local Similarity 100.0%; Pred. No. 2.43e+01;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 74151 ttcataatgaag 74163  
 ||||||||||||  
 Cp 13 TTCATTATGACG 1

Search completed: Sat Nov 27 15:14:18 1999  
 Job time : 51 secs.







CC MOLECULE TYPE: subgenomic DNA  
CC DESCRIPTION: /desc= "clone 1 insert"  
CC HYPOTHETICAL: NO  
CC ANTI-SENSE: NO  
CC ORIGINAL SOURCE:  
CC ORGANISM: Banana Bunchy Top Virus (BBTV)  
SQ SEQUENCE 287 BP; 98 A; 48 C; 76 G; 65 T; 0 OTHER.

Query Match 63.2%; Score 12; DB 3; Length 287;  
Best Local Similarity 100.0%; Pred. No. 1.15e+01;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 236 TTCATTAAATGAG 247  
CP 12 TTCATTAAATGAG 1

RESULT 2  
ID US-08-332-643-49 STANDARD; DNA; UNC; 730 BP.  
AC xxxxxx

DE Sequence 49, Application US/08332643  
CC Patent No. 5639634  
CC GENERAL INFORMATION:

CC APPLICANT: Suzuki, Shintaro  
CC TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS  
CC NUMBER OF SEQUENCES: 56  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
CC STREET: Two First National Plaza, 20 South Clark  
CC CITY: Chicago  
CC STATE: Illinois  
CC COUNTRY: USA  
CC ZIP: 60603

CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/332,643  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/872,643  
CC FILING DATE:

CC ATTORNEY/AGENT INFORMATION:  
CC NAME: No. 5639634and, Greta E.  
CC REGISTRATION NUMBER: 35,302  
CC REFERENCE/DOCKET NUMBER: 27866/30795  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (312) 346-5750  
CC TELEFAX: (312) 984-9740  
CC TELEX: 25-3856  
CC INFORMATION FOR SEQ ID NO: 49:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 730 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cDNA  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 2..730  
SQ SEQUENCE 730 BP; 209 A; 177 C; 164 G; 180 T; 0 OTHER.

Query Match 63.2%; Score 12; DB 1; Length 730;  
Best Local Similarity 100.0%; Pred. No. 1.15e+01;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 386 TTCATTAAATGAA 397  
QY 2 TTCATTAAATGAA 13

RESULT 3  
ID US-08-188-228-55 STANDARD; DNA; UNC; 730 BP.  
AC xxxxxx

DE Sequence 55, Application US/08188228  
CC Sequence 55, Application US/08188228  
CC Patent No. 5597725  
CC GENERAL INFORMATION:

CC APPLICANT: Suzuki, Shintaro  
CC TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS  
CC NUMBER OF SEQUENCES: 62  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
CC STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CC CITY: Chicago  
CC STATE: Illinois  
CC COUNTRY: USA  
CC ZIP: 60606

CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/188,228  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/049,460  
CC FILING DATE: 19 APR 1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/872,643  
CC FILING DATE: 17 APR 1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: No. 5597725and, Greta E.  
CC REGISTRATION NUMBER: 35,302  
CC REFERENCE/DOCKET NUMBER: 31340  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (312) 474-6300  
CC TELEFAX: (312) 474-0448  
CC TELEX: 25-3856  
CC INFORMATION FOR SEQ ID NO: 55:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 730 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cDNA  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 2..730  
SQ SEQUENCE 730 BP; 209 A; 177 C; 164 G; 180 T; 0 OTHER.

Query Match 63.2%; Score 12; DB 1; Length 730;  
Best Local Similarity 100.0%; Pred. No. 1.15e+01;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 386 TTCATTAAATGAA 397  
QY 2 TTCATTAAATGAA 13

RESULT 4  
ID US-08-332-638-55 STANDARD; DNA; UNC; 730 BP.  
AC xxxxxx  
DE Sequence 55, Application US/08332638



CC Sequence 55, Application US/08332638  
CC Patent No. 5646250  
CC  
CC GENERAL INFORMATION:  
CC APPLICANT: Suzuki, Shintaro  
CC TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS  
CC NUMBER OF SEQUENCES: 62  
CC CORRESPONDENCE ADDRES:  
CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
CC ADDRESSEE: Borun  
CC STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CC CITY: Chicago  
CC STATE: Illinois  
CC COUNTRY: USA  
CC ZIP: 60606  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/332,638  
CC FILING DATE: 01-NOV-1994  
CC CLASSIFICATION: 435  
CC  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/872,643  
CC FILING DATE: 17 APR 1992  
CC APPLICATION NUMBER: US/08/049,460  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: No. 5646250and, Greta E.  
CC REGISTRATION NUMBER: 35,302  
CC REFERENCE/DOCKET NUMBER: 31340  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (312) 474-6300  
CC TELEFAX: (312) 474-0448  
CC TELEX: 25-3856  
CC INFORMATION FOR SEQ ID NO: 55:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 730 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cDNA  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 2..730  
CC  
CC SEQUENCE 730 BP: 209 A; 177 C; 164 G; 180 T; 0 OTHER.

ID RESULT 5  
 US -08-188-228-55 STANDARD; DNA; UNC; 730 BP.  
 AC xxxxxx  
 DT  
 DE Sequence 55, Application US/08188228  
 CC Sequence 55, Application US/08188228  
 CC Patent No. 5597725  
 CC  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Suzuki, Shintaro  
 CC TITLE OF INVENTION: CACHERIN MATERIALS AND METHODS  
 CC  
 CC NUMBER OF SEQUENCES: 62  
 CC  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
 CC ADDRESSED: Botun  
 CC  
 CC STREET: 6300 Sears Tower, 233 S. Wacker Drive  
 CC CITY: Chicago

```

CC STATE: Illinois
CC COUNTRY: USA
CC ZIP: 60606
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/188,228
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/049,460
CC FILING DATE: 19 APR 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/872,643
CC FILING DATE: 17 APR 1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: NO. 5597725and, Greta E.
CC REGISTRATION NUMBER: 35,302
CC REFERENCE/DOCKET NUMBER: 31340
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (312) 474-6300
CC TELEFAX: (312) 474-0448
CC TELEX: 25-3856
CC INFORMATION FOR SEQ ID NO: 55:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 730 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA
CC FEATURE:
CC NAME/KEY: CDS
CC LOCATION: 2..730
CC SEQUENCE 730 BP: 209 A; 177 C; 164 G; 180 T; 0 OTHER.
SQ
Query Match 63.2%; Score 12; DB 1; Length 730;
Best Local Similarity 100.0%; Pred. No. 1.15e+01;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

Db 386 TTCATTAATGAA 397
CP 13 TTCATTAATGAA 2

RESULT 6
ID US-08-332-643-49 STANDARD; DNA; UNC; 730 BP.
AC xxxxxx
DE Sequence 49, Application US/08332643
CC Sequence 49, Application US/08332643
CC Patent No. 5639634
CC GENERAL INFORMATION:
CC APPLICANT: Suzuki, Shintaro
CC TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
CC NUMBER OF SEQUENCES: 56
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
CC ADDRESSEE: Bicknell
CC STREET: Two First National Plaza, 20 South Clark
CC STREET: Street
CC CITY: Chicago
CC STATE: Illinois
CC COUNTRY: USA
CC ZIP: 60603
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:

```

```

RESULT 5
ID US-08-332-643-49 STANDARD; DNA; UNC; 730 BP.
DT #####
PC
DE Sequence 49, Application US/08332643
CC Sequence 49, Application US/08332643
CC Patent No. 56396634
CC GENERAL INFORMATION:
CC APPLICANT: Suzuki, Shintaro
CC TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
CC NUMBER OF SEQUENCES: 56
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
CC ADDRESSEE: Bicknell
CC STREET: Two First National Plaza, 20 South Clark
CC STREET: Street
CC CITY: Chicago
CC STATE: Illinois
CC COUNTRY: USA
CC ZIP: 60603
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:

```



CC APPLICATION NUMBER: US/08/332,643  
 CC FILING DATE:  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US/07/872,643  
 CC FILING DATE:  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: No. 5639634and, Greta E.  
 CC REGISTRATION NUMBER: 35,302  
 CC REFERENCE/DOCKET NUMBER: 27866/30795  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (312) 346-5750  
 CC TELEFAX: (312) 984-9740  
 CC TELE: 25-3856  
 CC INFORMATION FOR SEQ ID NO: 49:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 730 base pairs  
 CC TYPE: nucleic acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: CDNA  
 CC FEATURE:  
 CC NAME/KEY: CDS  
 CC LOCATION: 2..730  
 CC SEQUENCE 730 BP; 209 A; 177 C; 164 G; 180 T; 0 OTHER.  
 SQ  
 Query Match 63.2%; Score 12; DB 1; Length 730;  
 Best Local Similarity 100.0%; Pred. No. 1.15e+01;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 386 TTCATTATGAA 397  
 |||||  
 Cp 13 TTCATTATGAA 2

RESULT 7  
 ID US-08-332-638-55 STANDARD; DNA; UNC; 730 BP.  
 AC xxxxxx  
 DE Sequence 55, Application US/08332638  
 CC Sequence 55, Application US/08332638  
 CC Patent No. 5646250  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Suzuki, Shintaro  
 CC TITLE OF INVENTION: CADCHEM MATERIALS AND METHODS  
 CC NUMBER OF SEQUENCES: 62  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
 CC STREET: 6300 Sears Tower, 233 S. Wacker Drive  
 CC CITY: Chicago  
 CC STATE: Illinois  
 CC COUNTRY: USA  
 CC ZIP: 60606  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: floppy disk  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/332,638  
 CC FILING DATE: 01-NOV-1994  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 07/872,643  
 CC FILING DATE: 17 APR 1992  
 CC APPLICATION NUMBER: US/08/049,460  
 CC FILING DATE:  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: No. 5646250and, Greta E.  
 CC REGISTRATION NUMBER: 35,302  
 CC REFERENCE/DOCKET NUMBER: 31340  
 CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (312) 474-6300  
 CC TELEFAX: (312) 474-0448  
 CC TELE: 25-3856  
 CC INFORMATION FOR SEQ ID NO: 55:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 730 base pairs  
 CC TYPE: nucleic acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: CDNA  
 CC FEATURE:  
 CC NAME/KEY: CDS  
 CC LOCATION: 2..730  
 CC SEQUENCE 730 BP; 209 A; 177 C; 164 G; 180 T; 0 OTHER.  
 SQ  
 Query Match 63.2%; Score 12; DB 1; Length 730;  
 Best Local Similarity 100.0%; Pred. No. 1.15e+01;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 386 TTCATTATGAA 397  
 |||||  
 Cp 13 TTCATTATGAA 2

RESULT 8  
 ID US-07-971-558-1 STANDARD; DNA; UNC; 1340 BP.  
 AC xxxxxx  
 DE Sequence 1, Application US/07971558  
 CC Sequence 1, Application US/07971558  
 CC Patent No. 5441736  
 CC GENERAL INFORMATION:  
 CC APPLICANT: GERLACH, GERALD F.  
 CC APPLICANT: WILSON, PHILIP J.  
 CC APPLICANT: ROSSI-CAMPOS, AMALIA  
 CC APPLICANT: POTTER, ANDREW A.  
 CC TITLE OF INVENTION: ACTINOBACILLUS PLEUROPNEUMONIAE OUTER  
 CC MEMBRANE LIPOPROTEIN A AND USES THEREOF  
 CC NUMBER OF SEQUENCES: 2  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: ROBERTA L. ROBINS  
 CC STREET: 635 BRYANT STREET  
 CC CITY: PALO ALTO  
 CC STATE: CALIFORNIA  
 CC COUNTRY: UNITED STATES OF AMERICA  
 CC ZIP: 94301  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: floppy disk  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/07/971,558  
 CC FILING DATE: 19921105  
 CC CLASSIFICATION: 435  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: ROBINS, ROBERTA L.  
 CC REGISTRATION NUMBER: 33,208  
 CC REFERENCE/DOCKET NUMBER: 9000-0027  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (415) 617-8999  
 CC TELEFAX: (415) 327-3231  
 CC INFORMATION FOR SEQ ID NO: 1:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 1340 base pairs  
 CC TYPE: NUCLEIC ACID  
 CC STRANDEDNESS: double  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: DNA (genomic)  
 CC FEATURE:  
 CC NAME/KEY: CDS  
 CC LOCATION: 158..1252  
 CC SEQUENCE 1340 BP; 502 A; 183 C; 265 G; 390 T; 0 OTHER.



Query Match 63.2%; Score 12; DB 1; Length 1340;  
Best Local Similarity 100.0%; Pred. No. 1.15e+01;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 826 AATGAACGATGC 837  
|||||  
QY 8 AATGAACGATGC 19

## RESULT 9

ID US-08-753-233-1 STANDARD; DNA; UNC; 1521 BP.  
AC xxxxxx

DE Sequence 1, Application US/08753233  
CC Sequence 1, Application US/08753233  
CC Patent No. 5728568  
CC GENERAL INFORMATION:  
CC APPLICANT: Sullivan, Francis  
CC APPLICANT: Kumar, Ravindra  
CC TITLE OF INVENTION: Human GDP-Mannose Hydratase  
CC NUMBER OF SEQUENCES: 5  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Genetics Institute, Inc.  
CC STREET: 87 Cambridgepark Drive  
CC CITY: Cambridge  
CC STATE: Massachusetts  
CC COUNTRY: USA  
CC ZIP: 02140  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/753,233  
CC FILING DATE:

CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Brown, Scott A.  
CC REGISTRATION NUMBER: 32,724  
CC REFERENCE/DOCKET NUMBER: G15285  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 498-8224  
CC TELEFAX: (617) 876-5851  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1521 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cDNA  
CC SEQUENCE 1521 BP; 357 A; 396 C; 429 G; 339 T; 0 OTHER.

Query Match 63.2%; Score 12; DB 2; Length 1521;  
Best Local Similarity 100.0%; Pred. No. 1.15e+01;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 363 TCATTAATGAAG 374  
|||||  
CP 12 TCATTAATGAAG 1

## RESULT 10

ID US-08-984-246-1 STANDARD; DNA; UNC; 1521 BP.  
AC xxxxxx

DE Sequence 1, Application US/08984246  
CC Sequence 1, Application US/08984246  
CC Patent No. 586307  
CC GENERAL INFORMATION:  
CC APPLICANT: Sullivan, Francis

CC APPLICANT: Kitz, Ron  
CC APPLICANT: Kumar, Ravindra  
CC TITLE OF INVENTION: Human GDP-Mannose Hydratase  
CC NUMBER OF SEQUENCES: 5  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Genetics Institute, Inc.  
CC STREET: 87 Cambridgepark Drive  
CC CITY: Cambridge  
CC STATE: Massachusetts  
CC COUNTRY: USA  
CC ZIP: 02140  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/984,246  
CC FILING DATE:  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/753,233  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Brown, Scott A.  
CC REGISTRATION NUMBER: 32,724  
CC REFERENCE/DOCKET NUMBER: G15285  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 498-8224  
CC TELEFAX: (617) 876-5851  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1521 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cDNA  
CC SEQUENCE 1521 BP; 357 A; 396 C; 429 G; 339 T; 0 OTHER.

Query Match 63.2%; Score 12; DB 3; Length 1521;  
Best Local Similarity 100.0%; Pred. No. 1.15e+01;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 363 TCATTAATGAAG 374  
|||||  
CP 12 TCATTAATGAAG 1

RESULT 11  
ID US-08-426-236-1 STANDARD; DNA; UNC; 3344 BP.  
AC xxxxxx

DE Sequence 1, Application US/08426236  
CC Sequence 1, Application US/08426236  
CC Patent No. 5629188  
CC GENERAL INFORMATION:  
CC APPLICANT: Shiba, Kiyotaka  
CC APPLICANT: Schimmel, Paul R.  
CC APPLICANT: Ripmaster, Tracy L.  
CC TITLE OF INVENTION: Human Alanyl-tRNA Synthetase Proteins,  
CC NUMBER OF SEQUENCES: 8  
CC Nucleic Acids and Tester Strains Comprising Same  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
CC STREET: Two Militia Drive  
CC CITY: Lexington  
CC STATE: Massachusetts  
CC COUNTRY: USA  
CC ZIP: 02173  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS



CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/426,236  
CC FILING DATE:  
CC CLASSIFICATION: 514  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Brook, David E.  
CC REGISTRATION NUMBER: 22,592  
CC REFERENCE/DOCKET NUMBER: CIP94-14  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 617-861-6240  
CC TELEFAX: 617-861-9540  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 344 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cDNA to mRNA  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 111..3014  
CC SEQUENCE 3344 BP: 827 A; 823 C; 966 G; 728 T; 0 OTHER.  
SQ  
Query Match 63.2%; Score 12; DB 1; Length 3344;  
Best Local Similarity 100.0%; Pred. No. 1.15e+01;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1231 TCATTATGAG 1242  
Cp 12 TCATTATGAG 1  
RESULT 12  
ID US-08-370-319C-12 STANDARD; DNA; UNC; 4129 BP.  
AC xxxxxx  
DE Sequence 12, Application US/08370319C  
CC Sequence 12, Application US/08370319C  
CC Patent No. 5856091  
CC GENERAL INFORMATION:  
CC APPLICANT: Brichard, Vincent; Van Pel, Aline;  
CC APPLICANT: Traversari, Catia; W lfel, Thomas; Coulie, Pierre;  
CC APPLICANT: Boon-Fallout, Thierry; De Pleen, Etienne  
CC TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A  
CC TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PERCURSOR PROCESSED TO AT LEAST ONE  
CC NUMBER OF SEQUENCES: 12  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Felfe & Lynch  
CC STREET: 805 Third Avenue  
CC CITY: New York City  
CC STATE: New York  
CC COUNTRY: USA  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
CC COMPUTER: IBM PS/2  
CC OPERATING SYSTEM: PC-DOS  
CC SOFTWARE: Wordperfect  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/370,319C  
CC FILING DATE: 10-JANUARY-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/272,351  
CC FILING DATE: 8-JULY-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/032,978  
CC FILING DATE: 18-MAR-1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Hanson, No. 5856091man D.  
CC REGISTRATION NUMBER: 30,946

CC REFERENCE/DOCKET NUMBER: LUD 5377.1  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 688-9200  
CC TELEFAX: (212) 838-3884  
CC INFORMATION FOR SEQ ID NO: 12:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 4129 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC FEATURE:  
CC OTHER INFORMATION: The sequence is preceded by an  
CC OTHER INFORMATION: unsequenced portion of from 4.7 to 5.3  
CC OTHER INFORMATION: kilobases  
SQ SEQUENCE 4129 BP: 1300 A; 827 C; 728 G; 1273 T; 1 OTHER.  
Query Match 63.2%; Score 12; DB 3; Length 4129;  
Best Local Similarity 100.0%; Pred. No. 1.15e+01;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 2713 GTTCATTATGA 2724  
Cp 14 GTTCATTATGA 3  
RESULT 13  
ID PCT-US94-01149-54 STANDARD; DNA; UNC; 7616 BP.  
AC xxxxxx  
DE Sequence 54, Application PC/TUS9401149  
CC Sequence 54, Application PC/TUS9401149  
CC GENERAL INFORMATION:  
CC APPLICANT: Shatzman, Allan  
CC APPLICANT: Scott, Miller  
CC APPLICANT: Dillon, Susan B.  
CC APPLICANT: Kane, James  
CC TITLE OF INVENTION: Vaccinal Polypeptides  
CC NUMBER OF SEQUENCES: 72  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: SmithKline Beecham Corporation - Corporate  
CC STREET: U. S. Mailcode UW2220 - 709 Swedeland Road  
CC CITY: King of Prussia  
CC STATE: Pennsylvania  
CC COUNTRY: USA  
CC ZIP: 19406-2799  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US94/01149  
CC FILING DATE:  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 149,150  
CC FILING DATE: 05-NOV-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 013,415  
CC FILING DATE: 01-FEB-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 108,914  
CC FILING DATE: 18-AUG-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 837,773  
CC FILING DATE: 18-FEB-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 751,896  
CC FILING DATE: 30-AUG-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 387,200  
CC FILING DATE: 28-JUL-1989



CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 238,801  
CC FILING DATE: 02-NOV-1988  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 645,732  
CC FILING DATE: 30-AUG-1984  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Baumeister, Kirk  
CC REGISTRATION NUMBER: 33,833  
CC REFERENCE/DOCKET NUMBER: P50134 PCT  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 215-270-5096  
CC TELEFAX: 215-270-5090  
CC INFORMATION FOR SEQ ID NO: 54:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 7616 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: double  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: DNA (genomic)  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 1879..2790  
CC SEQUENCE 7616 BP; 1913 A; 1926 C; 1936 G; 1841 T; 0 OTHER.  
SO  
Query Match 63.2%; Score 12; DB 4; Length 7616;  
Best Local Similarity 100.0%; Pred. No. 1.15e+01;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 1041 GCATCGTTCATT 1052  
CP 19 GCATCGTTCATT 8  
RESULT 14  
ID US-08-484-044-10 STANDARD; DNA; UNC; 11613 BP.  
AC xxxxxx  
DE Sequence 10, Application US/08484044  
CC Sequence 10, Application US/08484044  
CC Patent No. 5552282  
CC GENERAL INFORMATION:  
CC APPLICANT: Caskey, C. T.  
CC APPLICANT: Fu, Ying-Hui  
CC APPLICANT: Friedman, David L.  
CC APPLICANT: Pizutti, Antonio  
CC APPLICANT: Fenwick, Raymond G.  
CC TITLE OF INVENTION: Diagnosis of Myotonic Muscular Dystrophy  
CC NUMBER OF SEQUENCES: 13  
CC CORRESPONDENCE ADDRESSES:  
CC ADDRESSEE: Fulbright & Jaworski, L.L.P.  
CC STREET: 1301 McKinney, Suite 5100  
CC CITY: Houston  
CC STATE: Texas  
CC COUNTRY: U.S.A.  
CC ZIP: 77010-3095  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/484,044  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/019,940  
CC FILING DATE: 19-FEB-1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Paul, Thomas D.  
CC REGISTRATION NUMBER: 32,714  
CC REFERENCE/DOCKET NUMBER: D-5443  
CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 713/651-5325  
CC TELEFAX: 713/651-5246  
CC TELEX: 762829  
CC INFORMATION FOR SEQ ID NO: 10:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 11613 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC SEQUENCE 11613 BP; 2284 A; 3317 C; 3604 G; 2408 T; 0 OTHER.  
SO  
Query Match 63.2%; Score 12; DB 1; Length 11613;  
Best Local Similarity 100.0%; Pred. No. 1.15e+01;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 7536 TCATTATGAC 7547  
QY 3 TCATTATGAC 14  
RESULT 15  
ID US-08-374-983A-10 STANDARD; DNA; UNC; 872 BP.  
AC xxxxxx  
DE Sequence 10, Application US/08374983A  
CC Sequence 10, Application US/08374983A  
CC Patent No. 5631131  
CC GENERAL INFORMATION:  
CC APPLICANT: Jolivet, Jacques  
CC APPLICANT: Dayan, Alain  
CC APPLICANT: Bertrand, Richard  
CC TITLE OF INVENTION: cDNA Probes and Antibodies for Human  
CC TITLE OF INVENTION: Methylenetetrahydrofolate Synthetase  
CC NUMBER OF SEQUENCES: 14  
CC CORRESPONDENCE ADDRESSES:  
CC ADDRESSEE: Jacobson, Price, Holman & Stern  
CC STREET: 400 Seventh St. N.W.  
CC CITY: Washington D.C.  
CC COUNTRY: U.S.A.  
CC ZIP: 20004  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/374,983A  
CC FILING DATE: January 19, 1995  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Price, D. Douglas  
CC REGISTRATION NUMBER: 24,514  
CC REFERENCE/DOCKET NUMBER: 366/P57075A  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 202 638-6666  
CC TELEFAX: 202 393-5350  
CC TELEX: RCA 248593 IDEA UR  
CC INFORMATION FOR SEQ ID NO: 10:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 872 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC HYPOTHETICAL: YES  
CC ANTI-SENSE: NO  
SO SEQUENCE 872 BP; 270 A; 175 C; 216 G; 211 T; 0 OTHER.  
Query Match 57.9%; Score 11; DB 1; Length 872;  
Best Local Similarity 100.0%; Pred. No. 5.56e+01;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Mon Nov 29 13:18:42 1999


US-09-103-287-5.rni

Page 8

Db 717 GTTCATTAAAG 727  
|||||  
Cp 14 GTTCATTAAAG 4

Search completed: Sat Nov 27 15:15:11 1999  
Job time : 13 secs.







```

whole plants"
/db_xref="taxon:3702"
/clone_lib="NaCl-treated Arabidopsis subtraction library"
/dev_stage="7 day-old"
BASE COUNT      81 a      66 c      72 g      83 t      4 others
ORIGIN

Query Match      73.7%; Score 14; DB 12; Length 306;
Best Local Similarity 100.0%; Pred. No. 9.42e-03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      36 GTTCATTATGAAG 49
      14 GTTCATTATGAAG 1

RESULT 2
LOCUS      AA389840      316 bp      mRNA      EST      23-APR-1997
DEFINITION      OS210 NaCl-treated Arabidopsis subtraction library Arabidopsis
      thaliana cDNA 5' similar to Serine/threonine kinase NMK, mRNA
      sequence.
ACCESSION      AA389840
NID      G2042826
VERSION      AA389840.1 GI:2042826
KEYWORDS      EST.
SOURCE      thale cress.
ORGANISM      Arabidopsis thaliana
      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
      euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
      Rosidae; Caprales; Brassicaceae; Arabidopsis.
      1 (bases 1 to 316)
REFERENCE      Pih,K.T., Park,J.M., Jang,H.J., Kang,S.G., Piao,H.L. and Hwang,I.
      EST of salt inducible mRNA in Arabidopsis thaliana
      Unpublished (1997)
AUTHORS
      JOURNAL
      On Sep 12, 1996 this sequence version replaced gi:1394226.
COMMENT

Contact: Kyeong Tae Pih
Department of Plant Molecular Biology
Gyeongsang National Univ., Plant Molecular Biology and
Biotechnology Research Center
Jinju, Kyungnam 660-701, Korea
Tel: 82-591-751-5193
Fax: 82-591-759-9363
Email: ihwang@nongae.gsnu.ac.kr
Seq primer: T3
High quality sequence stop: 310.
FEATURES
      source
      location/Qualifiers
      1..316
      /organism="Arabidopsis thaliana"
      /strain="Columbia"
      /note="Subtracted cDNA library from salt(NaCl)-treated
      whole plants"
      /db_xref="taxon:3702"
      /clone_lib="NaCl-treated Arabidopsis subtraction library"
      /dev_stage="7 day-old"
BASE COUNT      86 a      66 c      77 g      83 t      4 others
ORIGIN

Query Match      73.7%; Score 14; DB 12; Length 316;
Best Local Similarity 100.0%; Pred. No. 9.42e-03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      36 GTTCATTATGAAG 49
      14 GTTCATTATGAAG 1

RESULT 3
LOCUS      AA389812      355 bp      mRNA      EST      23-APR-1997
DEFINITION      OS172 NaCl-treated Arabidopsis subtraction library Arabidopsis
      thaliana cDNA 5' mRNA sequence.
ACCESSION      AA389812
NID      92042798
VERSION
KEYWORDS
SOURCE
ORGANISM
      Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
      Primates; Catarrhini; Homnidae; Homo.
      1 (bases 1 to 355)
REFERENCE      Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
      Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
      Construction of a Characterized Clone Resource for Genomic
      Sequencing
      Unpublished (1998)
AUTHORS
      TITLE
      JOURNAL
      COMMENT
      Contact: Mahairas GG, Wallace JC, Hood L
      High Throughput Sequencing Center
      University of Washington
      401 Queen Anne Avenue North, Seattle, WA 98109, USA
      Tel: (206) 616-3618
      Fax: (206) 616-3887
      Email: jwallace@u.washington.edu
      Sequence Tagged Connector
      Plate: 3243 row: B column: 21

```

```

VERSION      AA389812.1 GI:2042798
KEYWORDS
SOURCE
ORGANISM
      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
      euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
      Rosidae; Caprales; Brassicaceae; Arabidopsis.
      1 (bases 1 to 355)
REFERENCE      Pih,K.T., Park,J.M., Jang,H.J., Kang,S.G., Piao,H.L. and Hwang,I.
      EST of salt inducible mRNA in Arabidopsis thaliana
      Unpublished (1997)
AUTHORS
      JOURNAL
      On Sep 12, 1996 this sequence version replaced gi:1394805.
COMMENT

Contact: Kyeong Tae Pih
Department of Plant Molecular Biology
Gyeongsang National Univ., Plant Molecular Biology and
Biotechnology Research Center
Jinju, Kyungnam 660-701, Korea
Tel: 82-591-751-5193
Fax: 82-591-759-9363
Email: ihwang@nongae.gsnu.ac.kr
Seq primer: T3
High quality sequence stop: 278.
FEATURES
      source
      location/Qualifiers
      1..355
      /organism="Arabidopsis thaliana"
      /strain="Columbia"
      /note="Subtracted cDNA library from salt(NaCl)-treated
      whole plants"
      /db_xref="taxon:3702"
      /clone_lib="NaCl-treated Arabidopsis subtraction library"
      /dev_stage="7 day-old"
BASE COUNT      98 a      80 c      80 g      93 t      4 others
ORIGIN

Query Match      73.7%; Score 14; DB 12; Length 355;
Best Local Similarity 100.0%; Pred. No. 9.42e-03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      36 GTTCATTATGAAG 49
      14 GTTCATTATGAAG 1

RESULT 4
LOCUS      AQ231915      367 bp      DNA      GSS      25-SEP-1998
DEFINITION      HS_3243_B1_A11_MR CIT Approved Human Genomic Sperm Library D Homo
      sapiens genomic clone Plate=3243 Col=21 Row=B, genomic survey
      sequence.
ACCESSION      AQ231915
NID      93657144
VERSION      AQ231915.1 GI:3657144
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
      Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
      Primates; Catarrhini; Homnidae; Homo.
      1 (bases 1 to 367)
REFERENCE      Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
      Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
      Construction of a Characterized Clone Resource for Genomic
      Sequencing
      Unpublished (1998)
AUTHORS
      TITLE
      JOURNAL
      COMMENT
      Contact: Mahairas GG, Wallace JC, Hood L
      High Throughput Sequencing Center
      University of Washington
      401 Queen Anne Avenue North, Seattle, WA 98109, USA
      Tel: (206) 616-3618
      Fax: (206) 616-3887
      Email: jwallace@u.washington.edu
      Sequence Tagged Connector
      Plate: 3243 row: B column: 21

```







Db 269 TCGTCATTATGA 282  
 Cp 16 TCGTCATTATGA 3

RESULT 7 A0132767 406 bp DNA GSS 22-SEP-1998  
 LOCUS HS\_3064\_B2\_H01\_MF C17 Approved Human Genomic Sperm Library D Homo  
 DEFINITION sapiens genomic clone Plate=3064 Col=2 Row=P, genomic survey  
 sequence.  
 ACCESSION A0132767  
 NID 93523493  
 VERSION A0132767.1 GI:3523493  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 406)  
 Authors: Mahairas,G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,  
 Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.F.  
 TITLE Construction of a Characterized Clone Resource for Genomic  
 Sequencing  
 JOURNAL Unpublished (1998)  
 COMMENT

CONTACT: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Sequence Tagged Connector  
 Plate: 3064 row: P column: 2  
 Class: BAC ends  
 High quality sequence stop: 406.

FEATURES  
 source  
 1..406  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in  
 E-Coli DH10B"  
 /db\_xref="taxon:9606"  
 /clone="Plate=3064 Col=2 Row=P"  
 /clone\_lib="C17 Approved Human Genomic Sperm Library D"  
 /sex="male"

BASE COUNT 103 a 84 c 85 g 134 t

ORIGIN  
 Query Match 73.7%; Score 14; DB 39; Length 406;  
 Best Local Similarity 100.0%; Pred. No. 9,42e-03;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 303 TCGTCATTATGA 316  
 Cp 16 TCGTCATTATGA 3

RESULT 8 R65414 406 bp mRNA EST 09-JAN-1998  
 LOCUS 13918 Lambda-PRL2 Arabidopsis thaliana cDNA clone 170N177, mRNA  
 DEFINITION  
 ACCESSION R65414  
 NID 9937881  
 VERSION R65414.1 GI:937881  
 KEYWORDS EST.  
 SOURCE Thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 Rosidae; Caprales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 406)  
 Authors: Newman,T., deRuijn,F.J., Green,P., Keegstra,K., Kende,H.,  
 McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M.,

TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT  
 On Apr 14, 1993 this sequence version replaced gi:838052.

CONTACT: Thomas Newman  
 MSU-DOE Plant Research Laboratory  
 Michigan State University  
 MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.  
 Lansing, MI  
 Tel: 517-353-0854  
 Fax: 517-353-9168  
 Email: 22313c@msu.edu  
 Seq primer: T7 dye primer.  
 Location/Qualifiers  
 1..406  
 /organism="Arabidopsis thaliana"  
 /strain="var columbia"  
 /note="Vector: lambda Zip-Lox; Site1: Sal; Site2: Not;  
 Lambda PRL2 is a cDNA library derived from equal  
 quantities of 4 pools of mRNA. The mRNA sources were 1) 7  
 day germinated etiolated seedlings; 2) tissue culture  
 grown roots; 3) staged plants half with 24 hour light  
 cycle, half on 16 hr light, 8 hour dark- rosettes; 4)  
 same plants as 3 but aerial tissue (stems, flowers and  
 siliques. The vector is BRL's lambda Zip-Lox. The cDNA  
 inserts were directionally cloned with Sal-Not arms using  
 oligo dt primed cDNA."  
 /db\_xref="taxon:3702"  
 /clone="170N177"  
 /clone\_lib="Lambda-PRL2"

BASE COUNT 98 a 87 c 95 g 110 t 16 others

ORIGIN  
 Query Match 73.7%; Score 14; DB 31; Length 406;  
 Best Local Similarity 100.0%; Pred. No. 9,42e-03;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 45 GTTCATTATGAG 58  
 Cp 14 GTTCATTATGAG 1

RESULT 9 A1017765 416 bp mRNA EST 27-AUG-1998  
 LOCUS OV32b07.X1 Soares-Lestis\_NHT Homo sapiens cDNA clone IMAGE:1638997  
 DEFINITION 3' similar to contains Alu repetitive element; contains element THR  
 repetitive element ;, mRNA sequence.  
 ACCESSION A1017765  
 NID 93232101  
 VERSION A1017765.1 GI:3232101  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 416)  
 Authors: NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT  
 On Oct 17, 1997 this sequence version replaced gi:2516451.

CONTACT: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be



found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.lnl.gov/db/imap/image.html

Insert length: 1540 Std Error: 0.00  
Seq primer: -40m3 fwd. ET from Amersham  
High quality sequence stop: 413.

## FEATURES

source

1. 416  
/organism="Homo sapiens"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from Clontech  
Laboratories, Inc., and primed with a Not I - oligo(dT)  
primer [5']  
TGTACCAATCTGAAGTGGAGCGCCGCAATTTTCTTTTCTTTT 3']  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization to Cot5, and was  
constructed by Bento Soares and M. Fatima Bonaldo."  
/db\_xref="taxon:9606"  
/map="640B03.1: 810E09; 17q21"  
/clone\_1ib="IMAGE:163897"  
/clone\_1ib="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/lab\_host="DH10B"

BASE COUNT

95 a 118 c 91 g 112 t

BASE COUNT

146 a 91 c 105 g 187 t 2 others

Query Match 73.7% Score 14; DB 21; Length 416;  
Best Local Similarity 100.0%; Pred. No. 9.42e-03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 73.7% Score 14; DB 27; Length 531;  
Best Local Similarity 100.0%; Pred. No. 9.42e-03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 169 CGTTCATTATGAA 182  
CP 15 CGTTCATTATGAA 2

Db 376 CTTCATTATGAA 389  
CY 1 CTTCATTATGAA 14

RESULT 10

LOCUS A1464854 531 bp mRNA EST 09-MAR-1999  
DEFINITION m292a02.y1 Soares mouse lymph node NBM1N Mus musculus cDNA clone

LOCUS A138575 550 bp mRNA EST 16-FEB-1997  
DEFINITION m08e10.r1 Soares mouse 3NDMS Mus musculus cDNA clone IMAGE:596874

ACCESSION

IMAG:720842 5, mRNA sequence.

ACCESSION

91700787

NID

A1464854

NID

A138575

VERSION

A1464854.1 GI:4318884

VERSION

A138575.1 GI:1700787

KEYWORDS

EST.

KEYWORDS

EST.

SOURCE

house mouse.

SOURCE

house mouse.

ORGANISM

Mus musculus

ORGANISM

Mus musculus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

1 (bases 1 to 531)  
Marras, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
Waterston, R. and Wilson, R.

AUTHORS

1 (bases 1 to 550)  
Marras, M., Hillier, L., Allen, M., Bowers, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

TITLE

The WashU-NCI Mouse EST Project 1999

TITLE

The WashU-HMI Mouse EST Project

JOURNAL

Unpublished (1999)

JOURNAL

Unpublished (1996)

COMMENT

On May 18, 1998 this sequence version replaced gi:3136517.

COMMENT

On Sep 12, 1996 this sequence version replaced gi:1394902.

CONTACT

Contact: Marra M/WashU-NCI Mouse EST Project 1999

CONTACT

Contact: Marra M/Mouse EST Project

WASHINGTON

Washington University School of Medicine

WASHINGTON

Washington University School of Medicine

4444

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

4444

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

TEL: 314

314 286 1800

TEL: 314

314 286 1800

FAX: 314

314 286 1810

FAX: 314

314 286 1810

EMAIL: mouseest@wustl.edu

Email: mouseest@wustl.edu

EMAIL: mouseest@wustl.edu

Email: mouseest@wustl.edu

THIS CLONE IS AVAILABLE ROYALTY-FREE THROUGH LNL; CONTACT THE

IMAG Consortium (info@image.lnl.gov) for further information.

THIS CLONE IS AVAILABLE ROYALTY-FREE THROUGH LNL; CONTACT THE

IMAG Consortium (info@image.lnl.gov) for further information.

THIS READ HAS BEEN VERIFIED (FOUND TO HIT ITS ORIGINAL SELF IN THE

correct orientation)

THIS READ HAS BEEN VERIFIED (FOUND TO HIT ITS ORIGINAL SELF IN THE

correct orientation)

PUTATIVE FULL LENGTH READ

vector to vector length is 655

PUTATIVE FULL LENGTH READ

vector to vector length is 655

FEATURES

source

FEATURES

source

1. 550

/organism="Mus musculus"

1. 550

/strain="C57BL/6J"



```

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCATCATGAGTGGAGCGCCGCTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Fatima Bonaldo."
/db_xref="taxon:10090"
/clone_image:596874"
/clone_11b="Soares mouse 3nBMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"

BASE COUNT      155 a      98 c      127 g      170 t
ORIGIN

Query Match      73.7%; Score 14; DB 36; Length 550;
Best Local Similarity 100.0%; Pred. No. 9.42e-03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      114 GTTCATTATGAG 127
      |||||
Cp      14 GTTCATTATGAG 1

RESULT 12
LOCUS      AF094894      1154 bp      DNA      GSS      28-FEB-1999
DEFINITION      Salmoneilla typhimurium clone 506-T7, Li-Cor strain LT2, genomic
survey sequence.
ACCESSION      AF094894
NID      94322736
VERSION      AF094894.1 GI:4322736
KEYWORDS      GSS
SOURCE      Salmoneilla typhimurium.
ORGANISM      Salmoneilla typhimurium
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmoneilla.
REFERENCE      1 (bases 1 to 1154)
AUTHORS      Wong,R.M.Y. and McClelland,M.
TITLE      End Sequences of Salmoneilla typhimurium LT2 Lambda DASH11 Clones,
Li-Cor
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 1154)
AUTHORS      Wong,R.M.Y. and McClelland,M.
TITLE      Direct Submission
JOURNAL      Submitted (28-SEP-1998) Molecular Biology, Sidney Kimmel Cancer
Center, 10835 Altman Row, San Diego, CA 92121, USA
FEATURES
source
1..1154
/organism="Salmoneilla typhimurium"
/strain="LT2"
/db_xref="taxon:602"
/note="end sequence from lambda clone"
/clone="506-T7, Li-Cor"
BASE COUNT      280 a      322 c      299 g      253 t
ORIGIN

Query Match      73.7%; Score 14; DB 41; Length 1154;
Best Local Similarity 100.0%; Pred. No. 9.42e-03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1019 TCGTCATTATGAG 1032
      |||||
Cp      16 TCGTCATTATGAG 3

RESULT 13
LOCUS      B95213      165 bp      DNA      GSS      25-JUN-1998
DEFINITION      CIT-HSP-2172J8.TF CIT-HSP Homo sapiens genomic clone 2172J8,

```

```

genomic survey sequence.
B95213
92977550
B95213.1 GI:2977550
GSS.
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 165)
AUTHORS      Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
Golden,M.K., Berry,K., Granger,D., Sun,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE      Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
JOURNAL      Unpublished (1997)
COMMENT      Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13-21;
Class: BAC ends.

FEATURES
source
1..165
/organism="Homo sapiens"
/note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
HindIII"
/db_xref="taxon:9606"
/clone="2172J8"
/clone_11b="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
BASE COUNT      53 a      28 c      35 g      49 t
ORIGIN

Query Match      68.4%; Score 13; DB 38; Length 165;
Best Local Similarity 100.0%; Pred. No. 2.66e-01;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      15 CGTCATTATGAG 27
      |||||
Cp      15 CGTCATTATGAG 3

RESULT 14
LOCUS      AA262179      176 bp      mRNA      EST      15-AUG-1997
DEFINITION      z925b04.s1 NCI-CCAP-GCBI Homo sapiens cDNA clone IMAGE:686191 3',
mRNA sequence.
ACCESSION      AA262179
NID      91898516
VERSION      AA262179.1 GI:1898516
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 176)
AUTHORS      NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      On Apr 14, 1993 this sequence version replaced gi:692762.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

```



This clone is available royalty-free through LNL; contact the IMAGE Consortium (infoimage.llnl.gov) for further information.  
 Insert length: 1370 Std Error: 0.00  
 Seq primer: -41ml3 fwd. ET from Amersham  
 High quality sequence stop: 150.  
 Location/Qualifiers

1. 1176  
 /organism="Homo sapiens"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+ IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTACCAATCTGAAGTGGAGCGCCGCTCATTTTCTTTTCTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:686191"  
 /clone\_lib="NCI CGAP\_GCB1"  
 /tissue\_type="germinal center B cell"  
 /lab\_host="DH10B"

BASE COUNT 56 a 38 c 33 g 49 t  
 ORIGIN  
 Query Match 68.4%; Score 13; DB 10; Length 176;  
 Best Local Similarity 100.0%; Pred. No. 2.66e-01;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 33 TTCATTATGAAG 45  
 |||||  
 Cp 13 TTCATTATGAAG 1

RESULT 15  
 LOCUS AA210523 217 bp mRNA EST 19-FEB-1997  
 DEFINITION mu30d06.r1 Soares 2NbMT Mus musculus cDNA clone IMAGE:640907 5',  
 mRNA sequence.  
 ACCESSION AA210523  
 NID g1807761  
 VERSION AA210523.1 GI:1807761  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 217)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 The WashU-HMI Mouse EST Project  
 Unpublished (1996)  
 On Sep 12, 1996 this sequence version replaced gi:1397482.

TITLE JOURNAL  
 COMMENT

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (infoimage.llnl.gov) for further information.  
 MGI:392899

Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 137.  
 Location/Qualifiers

FEATURES

source

1. 217  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTACCAATCTGAAGTGGAGCGCCGCTCATTTTCTTTTCTTTT-3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."  
 /db\_xref="taxon:10090"  
 /clone\_lib="IMAGE:640907"  
 /clone\_lib="Soares 2NbMT"  
 /sex="male"  
 /tissue\_type="Thymus"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"

BASE COUNT 74 a 25 c 37 g 81 t  
 ORIGIN  
 Query Match 68.4%; Score 13; DB 9; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 2.66e-01;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 148 TCGTTCATTATG 160  
 |||||  
 Cp 16 TCGTTCATTATG 4

Search completed: Sat Nov 27 15:13:06 1999  
 Job time : 236 secs.



**THIS PAGE BLANK (USPTO)**



\*\*\*\*\*  
 WILEY  
 (TM)  
 \*\*\*\*\*

Release 3.1a John F. Collins, Biocomputing Research Unit,  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MSrch\_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 15:15:30 1999; Maspar time 78.49 Seconds

Tabular output not generated. 670.987 Million cell updates/sec

Title: >US-09-103-287-6  
 Description: (1-19) from US09103287.seq  
 Perfect Score: 19  
 N.A. Sequence: 1 GTTACAAATATTAAAGAC 19  
 Comp: CAATGTTATTAATTTCTTC

Scoring table: TABLE Jmetric  
 Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database:

emb158  
 1:em\_ba1 2:em\_ba2 3:em\_fun 4:em\_htg 5:em\_hun1 6:em\_hun2  
 7:em\_in 8:em\_com 9:em\_or 10:em\_ov 11:em\_pat 12:em\_ph  
 13:em\_pl 14:em\_ro 15:em\_sts 16:em\_vl  
 genbank111

17:gb\_ba1 18:gb\_ba2 19:gb\_hgt1 20:gb\_hgt2 21:gb\_in1  
 22:gb\_in2 23:gb\_com 24:gb\_ov 25:gb\_pat 26:gb\_ph 27:gb\_pl1  
 28:gb\_pl2 29:gb\_pl1 30:gb\_pr2 31:gb\_pr3 32:gb\_ro  
 33:gb\_st 34:gb\_sts 35:gb\_sy 36:gb\_un 37:gb\_vl

Statistics: Mean 6.028; Variance 3.018; scale 1.997

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	100.0	1314	17	AF034076	Staphylococcus aureus	3.66e+01	
2	17	89.5	1044	29 HUMMCCPA04	Human mast cell carbox	6.10e+00	
3	17	89.5	96574	28 AT118B16	Arabidopsis thaliana D	6.10e+00	
4	16	84.2	1058	25 105615	Sequence 21 from Paten	2.36e+01	
5	16	84.2	80818	27 AB019232	Arabidopsis thaliana g	2.36e+01	
6	16	84.2	84107	29 HS248E1	Human DNA sequence fto	2.36e+01	
7	16	84.2	144659	31 AC006145	Homo sapiens PAC clone	2.36e+01	
8	16	84.2	155582	19 HS171N11	Human DNA sequence ***	2.36e+01	
9	16	84.2	173767	30 AC002449	Human PAC clone DJ404K	2.36e+01	
10	15	84.2	182606	19 HS288M22	Human DNA sequence ***	2.36e+01	
11	15	78.9	250	34 G14874	Human STS SHGC-13859	8.82e+01	
12	15	78.9	429	22 AF010534	Plasmodium falciparum	8.82e+01	
13	15	78.9	921	28 AF127454	Pemurus boldus rp116 ge	8.82e+01	

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
1	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	1	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
2	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	2	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
3	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	3	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
4	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	4	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
5	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	5	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
6	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	6	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
7	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	7	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
8	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	8	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
9	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	9	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
10	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	10	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
11	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	11	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
12	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	12	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
13	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	13	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
14	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	14	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
15	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	15	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
16	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	16	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
17	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	17	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
18	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	18	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
19	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	19	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
20	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	20	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
21	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	21	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
22	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	22	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
23	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	23	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
24	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	24	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
25	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	25	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
26	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	26	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
27	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	27	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
28	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	28	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
29	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	29	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
30	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	30	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
31	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	31	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
32	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	32	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
33	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	33	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
34	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	34	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
35	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	35	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
36	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	36	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
37	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	37	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
38	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	38	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
39	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	39	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
40	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	40	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
41	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	41	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
42	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	42	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
43	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	43	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
44	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	44	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers
45	AF034076	1314 bp DNA (murC) gene, complete cds.	AF034076	2642658	GI:2642658	Staphylococcus aureus	Staphylococcus aureus	45	Lowe, A.M. and Deresiewicz, R.L.	Involved in cell wall biosynthesis	Unpublished	Location/Qualifiers

#### ALIGNMENTS

RESULT 1  
 LOCUS AF034076 1314 bp DNA  
 DEFINITION Staphylococcus aureus UDP-N-acetylmuramoyl-L-alanine synthetase  
 ACCESSION AF034076  
 VERSION 2642658  
 KEYWORDS murC, gene, complete cds.  
 SOURCE AF034076.1 GI:2642658  
 ORGANISM Staphylococcus aureus.  
 Staphylococcus aureus.  
 Eubacteria; Firmicutes; Low G+C gram-positive bacteria;  
 Bacillaceae; Staphylococcus.  
 1 (bases 1 to 1314)  
 Lowe, A.M. and Deresiewicz, R.L.  
 Cloning and sequencing of Staphylococcus aureus murC, a gene  
 involved in cell wall biosynthesis  
 Unpublished  
 2 (bases 1 to 1314)  
 Lowe, A.M. and Deresiewicz, R.L.  
 Direct Submission  
 Submitted (11-NOV-1997) Channing Laboratory, Brigham and Women's  
 Hospital and Harvard Medical School, 181 Longwood Ave, Boston, MA  
 02115, USA  
 Location/Qualifiers  
 1..1314  
 /organism="Staphylococcus aureus"  
 /db\_xref="taxon:1280"  
 1..1314  
 /gene="murC"  
 1..1314  
 /gene="murC"  
 /note="murC"  
 /function="cell wall biosynthesis"  
 /note="murC: UDP-N-acetylmuramate-alanine ligase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="UDP-N-acetylmuramoyl-L-alanine synthetase"  
 /protein\_id="AAB87090.1"



```

/db_xref="PID:g2642659"
/db_xref="GI:2642659"
/translation="MTHYFVINGKSGNSLAQIMHDLGHEVGSDIENYFTEVALR
NKGIILPFDANNIKEDYVLOGNAPFASHEKRAHOMKLDVYVNDPLGQIDQY
SVAVGAGKSTTGLSHVMGDKSTFLIGDGMGLPESDYAFAPACEYRRHFLS
YKPDYATNTNIDFHPDVKNDVDFDAFOEAAHYKKGITAMGDEHLKIEANVP
YYGFKSDSDIYQNIQITDKGTADVDGEFDHFLSPQIGDHTLNLALVATIST
LEKLDVNIKELTEFGVKKRFRNETTIANOVYDVAHHPREISATLETARKKYPK
EVAVFQPHTESRTQAFNFEAFSLSKADRVFLCEIFGSIKRENTGALTIOIDLKIEG
ASLIMEDSINVLEFDNMAVLEFKGAGDIQKQNLQNAVLKLGKMKNAF"
BASE COUNT      467 a      163 c      250 g      434 t
ORIGIN
Query Match      100.0%; Score 19; DB 17; Length 1314;
Best Local Similarity 100.0%; Pred. No. 3.66e-01;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB      844 GTTACCAATATTAAAGAA 862
      1 GTTACCAATATTAAAGAA 19
OY

RESULT 2
LOCUS      HUMMCCPA04      1044 bp      DNA      PRI      07-JAN-1995
DEFINITION Human mast cell carboxypeptidase A (MC-CPA) gene, exon 10.
ACCESSION      M73719
VERSION      M73719.1 GI:187439
KEYWORDS      carboxypeptidase; metalloexopeptidase; secretory granules.
SEGMENT      4 of 5
SOURCE      Homo sapiens (tissue library: lambda-gHMCCPA-4) DNA.
ORGANISM      Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      Reynolds,D.S., Gurley,D.S. and Austen,K.F.
AUTHORS      Cloning and characterization of the novel gene for mast cell
TITLE      carboxypeptidase A
JOURNAL      J. Clin. Invest. 89 (1), 273-282 (1992)
MEDLINE      92105393
FEATURES
     source          1..1044
                     location/Qualifiers
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="mast"
                     /tissue_1id="lambda-gHMCCPA-1"
                     /tissue_1id="lambda-gHMCCPA-3"
                     /tissue_1id="lambda-gHMCCPA-4"
                     /map="unassigned"
                     order(M73718:1:5957..6192:1..466)
     intron          /gene="MC-CPA"
                     /number=9
     exon            /number=9
                     /gene="CPA3"
                     /note="G00-125-231"
                     /number=10
BASE COUNT      306 a      218 c      182 g      338 t
ORIGIN
Query Match      89.5%; Score 17; DB 29; Length 1044;
Best Local Similarity 100.0%; Pred. No. 6.10e+00;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB      598 TTACAATATTAAAGAA 614
      2 TTACAATATTAAAGAA 18
OY

RESULT 3
LOCUS      ATT18B16      96574 bp      DNA      PLN      25-MAR-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone T18B16 (ESSAILI
project).
ACCESSION      AL021687

```

```

NID      92828278
VERSION      AL021687.1 GI:2828278
KEYWORDS      thale cress.
SOURCE      Arabidopsis thaliana
ORGANISM      Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Caprariales; Brassicaceae; Arabidopsis.
REFERENCE      1 (bases 1 to 96574)
AUTHORS      Bevan,M., Benes,V., Rechmann,S., Borokova,D., Ansgorge,W.,
Bancroft,I., Mewes,H.W., Mayer,K. and Schueller,C.
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 96574)
AUTHORS      EU Arabidopsis sequencing project.
TITLE      Direct Submission
JOURNAL      Submitted (09-APR-1998) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bsrc.ac.uk
FEATURES
     source          1..96574
                     location/Qualifiers
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     /gene="T18B16.10"
                     /gene="T18B16.10"
                     /complement(join(956..1188,1311..1466,1728..1792,
2086..2181,2470..2570,2674..2714,3024..3062,3276..3391,
3835..3938,4063..4128,4231..4284,4460..4534,4660..4797))
                     /gene="T18B16.10"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CA16693.1"
                     /db_xref="PID:e1248929"
                     /db_xref="PID:g2828279"
                     /db_xref="GI:2828279"
                     /db_xref="SPTREMBL:O49667"
                     /translation="MSKRVYEGVMVRYGRKRGSYTHMRFYLEPRLAYKKKPOD
YQVPIKTMIDGNCRVEDRGKTHGHMVVLSYNNKRSKSHRTMAAFNIOALMKR
EKISVYDQHOESVPRNGQOVSFEYKSGMDTGTAASSSHESHLITCYCKRSAS
FRISADEEDSRSLKRRRTTIGNGALEDAISFMOSSLTFEHLVVDLPSCSA
MKAVGVAVATCEIEFELLMGMDGTRYMDSFGFSLVEVDGHTAVLYHRLLDWPP
MTWVRDLCYRVYRRNDDSGYGGYVNSLPKNGRPRVOVHLLIDIKKGAGYL
PAFOCHCLLOMLNSVAGLRWFESQTDRCGVHTRIPVYVNNASSSLSTKSGSLHKA
PSLDQTSVNRNSLIMDESDDDDEFQIAESEQV"
                     /complement(956..1188)
                     /gene="T18B16.10"
                     /number=1
                     /complement(1189..1310)
                     /gene="T18B16.10"
                     /number=1
                     /complement(1311..1466)
                     /gene="T18B16.10"
                     /number=2
                     /complement(1467..1727)
                     /gene="T18B16.10"
                     /number=2
                     /complement(1728..1792)
                     /gene="T18B16.10"
                     /number=3
                     /complement(1793..2085)
                     /gene="T18B16.10"
                     /number=3
                     /complement(2086..2181)
                     /gene="T18B16.10"
                     /number=4
                     /complement(2182..2469)
                     /gene="T18B16.10"
                     /number=4
                     /complement(2470..2570)
                     /gene="T18B16.10"

```



intron complement(2571. .2673)  
 /gene="T18B16.10"  
 /number=5  
 exon complement(2674. .2714)  
 /gene="T18B16.10"  
 /number=6  
 intron complement(2715. .3023)  
 /gene="T18B16.10"  
 /number=6  
 exon complement(3024. .3062)  
 /gene="T18B16.10"  
 /number=7  
 intron complement(3063. .3275)  
 /gene="T18B16.10"  
 /number=7  
 exon complement(3276. .3391)  
 /gene="T18B16.10"  
 /number=8  
 intron complement(3392. .3834)  
 /gene="T18B16.10"  
 /number=8  
 exon complement(3835. .3938)  
 /gene="T18B16.10"  
 /number=9  
 intron complement(3939. .4062)  
 /gene="T18B16.10"  
 /number=9  
 exon complement(4063. .4128)  
 /gene="T18B16.10"  
 /number=10  
 intron complement(4129. .4230)  
 /gene="T18B16.10"  
 /number=10  
 exon complement(4231. .4284)  
 /gene="T18B16.10"  
 /number=11  
 intron complement(4285. .4459)  
 /gene="T18B16.10"  
 /number=11  
 exon complement(4460. .4534)  
 /gene="T18B16.10"  
 /number=12  
 intron complement(4535. .4659)  
 /gene="T18B16.10"  
 /number=12  
 exon complement(4660. .4797)  
 /gene="T18B16.10"  
 /number=13  
 intron complement(5838. .5912)  
 /gene="T18B16.20"  
 /number=1  
 cds 5838. .11411  
 /gene="T18B16.20"  
 complement(join(5838. .5912,6020. .6183,6265. .6398,  
 6638. .6731,6805. .6913,7005. .7053,7174. .7190,7281. .7307  
 7733. .7790,7870. .8681,8733. .11411))  
 /gene="T18B16.20"  
 /note="similarity to hypothetical protein Y1106w,  
 Saccharomyces cerevisiae, PIR:S48466; Contains  
 ATP/GTP-binding site motif A (P-loop), pos.38-46; contains  
 EST gp:H7164, T04732"  
 /codon\_start=1  
 /translation="MEKODOTSREELIKRIMDSLAGDGVPSKTVLNGEAGIGKTWLAK  
 EVSQRTQGEVYNYLMHLINKKIDEDKSYITILAAQSLIYEFEGEGEPDLDYPLESL  
 KEIKIKEMIHKDKDNLLIIDDGSMTEEDVQGLNLOFLKEYSAVKLLIVRDER  
 EEKEESTIKHPTKEESLDLDAEDLITSFSEDMVPYLKRIKCNKLEKPTLMSCC

[illegible]



AUTHORS Rusche, J.R., Putney, S.D., Javaherian, K., Farley, J., Grimalta, R., Lynn, D., Petro, J., and O'Keefe, T.  
 TITLE Novel HIV proteins and peptides useful in the diagnosis, prophylaxis or therapy of AIDS  
 JOURNAL Patent: EP 0306219-A2 21 08-MAR-1989;  
 FEATURES Location/Qualifiers  
 source 1..1058 /organism="unknown"  
 BASE COUNT 341 a 188 c 188 g 341 t  
 ORIGIN

Query Match 84.2%; Score 16; DB 25; Length 1058;  
 Best Local Similarity 100.0%; Pred. No. 2.36e+01;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 815 TTCCTTAATTTGTA 830  
 ||||||||||||  
 Cp 18 TTCCTTAATTTGTA 3

RESULT 5  
 LOCUS AB019232 80818 bp DNA PLN 06-FEB-1999  
 DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, Pl clone: ML23, complete sequence.  
 ACCESSION AB019232  
 NID 93869071  
 VERSION AB019232.1 GI:3869071  
 KEYWORDS HTG.  
 SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone\_lib:Mitsui P1 clone:ML23.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Caprales; Brassicaceae; Arabidopsis.  
 REFERENCE 1 (bases 1 to 80818)  
 AUTHORS Nakamura, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-OCT-1998) to the DDBJ/EMBL/GenBank databases. Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of Gene Structure 2, 1532-3, Yana, Kisarazu, Chiba 292, Japan (E-mail:yasukazu@kazusa.or.jp, Tel:+81-438-52-3935, Fax:+81-438-52-3934)  
 FEATURES 2 (sites)  
 source 1..80818  
 Location/Qualifiers  
 1..80818 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="3"  
 /clone="ML23"  
 /clone\_lib="Mitsui P1"  
 BASE COUNT 26267 a 14707 c 14502 g 25342 t  
 ORIGIN

Query Match 84.2%; Score 16; DB 27; Length 80818;  
 Best Local Similarity 100.0%; Pred. No. 2.36e+01;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 51176 ACAATATTAAAGAG 51191  
 ||||||||||||  
 Oy 4 ACAATATTAAAGAG 19

RESULT 6  
 LOCUS HS248E1 84107 bp DNA PRI 24-JUL-1998  
 DEFINITION Human DNA sequence from clone 248E1 on chromosome 6q23.1-23.3 contains DOPAMINE-BETA-MONOOXYGENASE PRECURSOR, EF-1-ALPHA-2 pseudogene EST GSS and CA repeat, complete sequence.  
 ACCESSION AL023578  
 NID 93228200

VERSION AL023578.1 GI:3228200  
 KEYWORDS HTG; repeat polymorphism.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 84107)  
 AUTHORS Tubby, B.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-JUL-1998) E-mail enquiries: humquery@sanger.ac.uk  
 COMMENT On Jun 16, 1998 this sequence version replaced gi:3183866. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence is the entire insert of clone 248E1. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>  
 248E1 is from the library RPEC11 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:PCFAC2>.

FEATURES  
 source 1..84107  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone="248E1"  
 /map="q23.1-23.3"  
 /clone\_lib="RPEC11"  
 complement(52..405)  
 /note="ML1A1 repeat: matches 365. 1 of consensus"  
 546..845  
 /note="AluY repeat: matches 2. 301 of consensus"  
 1026..1051  
 /note="13 copies 2 mer aa 92% conserved"  
 1308..1622  
 /note="AluSg repeat: matches 1. 300 of consensus"  
 3308..3428  
 /note="MIR repeat: matches 84. 201 of consensus"  
 3550..3629  
 /note="2 copies 40 mer 89% conserved"  
 complement(5748..6186)  
 /note="MIR1C repeat: matches 466. 1 of consensus"  
 complement(8135..8435)  
 /note="AluY repeat: matches 301. 1 of consensus"  
 9695..9918  
 /note="MIR repeat: matches 21. 262 of consensus"  
 complement(10955..11329)  
 /note="MIR1B repeat: matches 337. 1 of consensus"  
 complement(11062..11329)  
 /note="MIR1A repeat: matches 274. 1 of consensus"  
 11239..11623  
 /note="match: GSS B93863 clone 2170K12"  
 12035..12514  
 /note="L1PA2 repeat: matches 418. 891 of consensus"  
 12896..18811  
 /gene="dJ248E1.1"  
 join(<12896..13075,13375..13477,17393..17559,18620..>18811)  
 /gene="dJ248E1.1"  
 /note="match: protein P09172"  
 /codon\_start=1  
 /evidence=not\_experimental



```

/product="d3248E1.1(DOPAMINE-BETA-MONOOXYGENASE PRECURSOR
(EC 1.14.17.1) (DOPAMINE BETA-HYDROXYLASE) (DBH))"
/protein_id="CA119040.1"
/db_xref="PID:e1312801"
/db_xref="PID:g3355516"
/db_xref="GI:3355516"
/db_xref="SPTRMBL:075647"
/translation="VEPTIQREHSELYHHLLIYOCNNFNDSVLESGECHYRPPPPD
FLTCTVIFPAAIIGGEFSYPYHVLSTGTPDPRVYLLEVHNDPTELEELINSGI
RLFTMDIPLKRDAGVIEAGLVWLSFTPLTPGPPEKSGHGTDLCELEALAEKPSG
HYFAVLILAHNLGRCIRLRHFRKGEKMLIAYDDDFDNFQFOYLRKEQITLP"
repeat__region
/note="MIR2 repeat: matches 146. .4 of consensus"
complement(15661. .15607)
/note="MIR repeat: matches 257. .97 of consensus"
complement(15810. .16171)
/note="MIR1 repeat: matches 365. .1 of consensus"
complement(16113. .16253)
/note="MIR repeat: matches 98. .18 of consensus"
complement(16371. .16623)
/misc_feature
/gene="d3248E1.1"
/note="match: 223591 STS containing (CA) repeat"
16503. .16535
/note="16 copies 2 mer g 100% conserved; differs from
223591"
repeat__region
16786. .16873
/note="MIR2 repeat: matches 57. .143 of consensus"
18189. .18273
/note="MIR repeat: matches 71. .151 of consensus"
18699. .25836
/note="match: 5' EST D59397 clone GEN-027G03"
complement(20216. .20503)
/note="AlusJ repeat: matches 291. .3 of consensus"
22923. .23570
/note="L1M3 repeat: matches 4. .685 of consensus"
25765. .44635
/note="match: multiple ESTs; match: M8407 AA424680 N995999
C01809 AA715255. .27858)
complement(27556. .27858)
/note="AlusJ repeat: matches 302. .1 of consensus"
complement(28723. .28872)
/note="match: 3' EST N67273 clone 286400"
28865. .28898
/note="17 copies 2 mer aa 85% conserved"
complement(30532. .31375)
/note="L1P10 repeat: matches 890. .1 of consensus"
complement(31234. .36025)
/note="L1 repeat: matches 5390. .631 of consensus"
complement(36534. .36770)
/note="L1M2 repeat: matches 351. .110 of consensus"
complement(37079. .37437)
/note="MIR1 repeat: matches 363. .4 of consensus"
complement(37820. .38085)
/note="MIR2 repeat: matches 327. .49 of consensus"
complement(38118. .38221)
/note="L1M3 repeat: matches 113. .1 of consensus"
complement(38408. .39028)
/note="L1 repeat: matches 5021. .4387 of consensus"
41577. .41608
/note="16 copies 2 mer ag 97% conserved"
complement(42052. .42871)
/note="L1P15 repeat: matches 904. .91 of consensus"
43170. .43302
/note="AlusJ repeat: matches 1. .132 of consensus"
44697. .45435
/note="match: multiple ESTs; match: AA604597 AA994351
W46980 N98655 AA530889; match: T17142 R32545 AA235998
AA618502; match: AA907610 AA233672 W47081
AA194117; match: AA969575 AA928115 AA194174 AA424574;
match: N26240 AA227906 AA321213"
45844. .45919
/note="2 copies 38 mer 91% conserved"
46810. .46866
repeat__region

```

```

repeat_region /note="3 copies 19 mer 84% conserved"
46880..46977 /note="MER42c repeat: matches 1436..1538 of consensus"
repeat_region /note="2 copies 51 mer 85% conserved"
47400..47501 /note="2 copies 51 mer 85% conserved"
repeat_region /note="MIR2 repeat: matches 146..68 of consensus"
/note="MIR2 repeat: matches 146..68 of consensus"
repeat_region /note="MIR repeat: matches 124..208 of consensus"
/note="MIR repeat: matches 124..208 of consensus"
misc_feature /note="match: 279094 chromosome 6 HindIII fragment"
53392..55012 /note="MIR2 repeat: matches 165..798 of consensus"
/note="MIR2 repeat: matches 165..798 of consensus"
repeat_region /note="AluYo repeat: matches 302..2 of consensus"
55798..56085 /note="MIR2 repeat: matches 797..1079 of consensus"
/note="MIR2 repeat: matches 797..1079 of consensus"
misc_feature /note="match: 279076 chromosome 6 HindIII fragment"
59224..59349 /note="MIR2 repeat: matches 21..146 of consensus"
/note="MIR2 repeat: matches 21..146 of consensus"
repeat_region /note="MIR5A repeat: matches 40..103 of consensus"
60284..60460 /note="AluY repeat: matches 118..297 of consensus"
/note="AluY repeat: matches 118..297 of consensus"
repeat_region /note="AluSg repeat: matches 303..1 of consensus"
62847..63119 /note="L1M4S repeat: matches 873..1045 of consensus"
/note="L1M4S repeat: matches 873..1045 of consensus"
repeat_region /note="AluSx repeat: matches 1..282 of consensus"
63214..63493 /note="THE1B repeat: matches 3..364 of consensus"
64112..64462 /note="MIR repeat: matches 190..97 of consensus"
/note="MIR repeat: matches 190..97 of consensus"
repeat_region /note="MIR repeat: matches 253..89 of consensus"
/note="MIR repeat: matches 253..89 of consensus"
repeat_region /note="MIR repeat: matches 140..25 of consensus"
67867..68129 /note="AluDb repeat: matches 40..300 of consensus"
/note="AluDb repeat: matches 40..300 of consensus"
gene /gene="dJ248E1.2"
complement(68162..69508)
CDS /gene="dJ248E1.2"
/codon_start=1

Note: remainder of annotations omitted.

Query Match 84.2% Score 16; DB 29; Length 84107;
Best Local Similarity 100.0%; Pred. No. 2.3e+01;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 49645 CTCTCTTAATATTGT 49660
|||||
Cp 19 CTCTTAATATTGT 4

RESULT 7
LOCUS AC006145 144659 bp DNA PRI 18-MAR-1999
DEFINITION Homo sapiens PAC clone DJ0560014 from 7q21.1-q21.2, complete
ACCESSION AC006145
NID g4454525
VERSION AC006145.2 GI:4454525
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 144659)
DIFFERENCE

```



AUTHORS Sulston, J.E. and Waterston, R.  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 99063792  
REFERENCE 2 (bases 1 to 144659)  
AUTHORS Mead, R. and Bauer, C.  
TITLE The sequence of Homo sapiens PAC clone DJ0560014  
JOURNAL Unpublished (1999)  
REFERENCE 3 (bases 1 to 144659)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (07-DEC-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
4 (bases 1 to 144659)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (18-MAR-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Mar 19, 1999 this sequence version replaced g1:3980538.  
SUBMITTED BY: WUCSC  
Genome Sequencing Center  
Department of Genetics  
Washington University  
St. Louis MO 63108, USA  
http://genome.wustl.edu/gsc  
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see  
http://www.nhgri.nih.gov/DIR/GRB/CHR7, send  
mailto:sgreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:  
This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://pacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.  
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.  
VECTOR: pCYPAC2  
NEIGHBORING SEQUENCE INFORMATION:  
Actual start of this clone is at base position 1 of DJ0560014; actual end is at 144659 of DJ0560014

A run of A's extends from base 112507 to 112528 of DJ0560014. The data through this region differs in the number of A's represented. The submitted sequence reflects the longest run of A's in any one subclone.

FEATURES  
source location/Qualifiers  
1..144659  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/chromosome="7"  
/map="7q21.1-q21.2"  
/clone\_lib="RPCI-4"  
/clone="DJ0560014"  
repeat\_region 377..2556  
/rpt\_family="L1"  
repeat\_region 3789..3837  
/rpt\_family="AT\_rich"  
repeat\_region 4406..4427  
/rpt\_family="AT\_rich"  
repeat\_region 4634..4690  
/rpt\_family="AT\_rich"  
repeat\_region 4864..5227  
/rpt\_family="L1"  
repeat\_region 5228..5251  
/rpt\_family="(CA)n"  
repeat\_region 5251..5273  
/rpt\_family="(TA)n"  
repeat\_region 5252..7075  
/rpt\_family="L1"  
repeat\_region 7083..8086  
/rpt\_family="L1"  
repeat\_region 8087..8125  
/rpt\_family="(TAA)n"  
repeat\_region 8126..8406  
/rpt\_family="Alu"  
repeat\_region 8407..11673  
/rpt\_family="L1"  
repeat\_region 11674..12133  
/rpt\_family="Retroviral"  
repeat\_region 12134..14212  
/rpt\_family="L1"  
repeat\_region 14213..14673  
/rpt\_family="MER4-group"  
repeat\_region 14677..14878  
/rpt\_family="L1"  
repeat\_region 14890..15211  
/rpt\_family="MAlR"  
repeat\_region 15212..16375  
/rpt\_family="L1"  
repeat\_region 16386..16406  
/rpt\_family="(CA)n"  
repeat\_region 16879..17138  
/rpt\_family="L1"  
repeat\_region 17139..17501  
/rpt\_family="MAlR"  
repeat\_region 17502..17593  
/rpt\_family="L1"  
repeat\_region 17595..17965  
/rpt\_family="MAlR"  
repeat\_region 17966..18707  
/rpt\_family="MAlR"  
repeat\_region 18704..19073  
/rpt\_family="L1"  
repeat\_region 19074..19115  
/rpt\_family="(CA)n"  
repeat\_region 19116..19722  
/rpt\_family="L1"  
repeat\_region 19736..20031  
/rpt\_family="Alu"  
repeat\_region 20041..20489  
/rpt\_family="L1"  
repeat\_region 20629..20660  
/rpt\_family="AT\_rich"  
repeat\_region 20661..21005  
/rpt\_family="L1"  
repeat\_region 21008..21067  
/rpt\_family="AT\_rich"  
repeat\_region 21092..21598  
/rpt\_family="L1"  
repeat\_region 21618..21778  
/rpt\_family="L1"  
repeat\_region 21779..22950  
/rpt\_family="L1"



```

Repeat_region      /rpt_family="L1"
                    22941. .24161
Repeat_region      /rpt_family="L1"
                    24173. .23137
Repeat_region      /rpt_family="L1"
                    25226. .25301
Repeat_region      /rpt_family="L1"
                    25301. .25493
Repeat_region      /rpt_family="L1"
                    25531. .25638
Repeat_region      /rpt_family="L1"
                    25670. .25925
Repeat_region      /rpt_family="L1"
                    25939. .26064
Repeat_region      /rpt_family="Alu"
                    26065. .26713
Repeat_region      /rpt_family="L1"
                    26714. .27006
Repeat_region      /rpt_family="Alu"
                    27007. .27240
Repeat_region      /rpt_family="L1"
                    27241. .27618
Repeat_region      /rpt_family="L1"
                    27619. .29417
Repeat_region      /rpt_family="L1"
                    29999. .31616
Repeat_region      /rpt_family="MER4_group"
                    31925. .31948
Repeat_region      /rpt_family="AT_rich"
                    32785. .32855
Repeat_region      /rpt_family="purine-rich"
                    32857. .32916
Repeat_region      /rpt_family="L1"
                    35865. .35915
Repeat_region      /rpt_family="MER1_type"
                    35926. .36325
Repeat_region      /rpt_family="Alu"
                    36528. .36637
Repeat_region      /rpt_family="MR"
                    37639. .37802
Repeat_region      /rpt_family="Alu"
                    38039. .38080

```

Note: remainder of annotations omitted

```

Query Match      84.2%;   Score 16; DB 31; Length 144659;
Best Local Similarity 100.0%; Pred. No. 2,366+0;
Matches    16; Conservative    0; Mismatches    0; Indels    0; Gaps    0

```

RESULT	LOCUS	DEFINITION	ACCESSION	NID	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
8	HS171N11	Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 171N11, WORKING DRAFT SEQUENCE.	155582 bp	DNA	HTG	11-MAR-1999							
	AL031433		94481851	AL031433.2	GI:481851	HTG; HTGS_PHASE1.	human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	1 (Bases 1 to 155582)	Phillips,S.	Direct Submission	Submitted (09-MAR-1999) Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires: humuquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Mar 23, 1999 this sequence version replaced gi14375918.

```

FEATURES
source

IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Unfinished: djl71n11 Contig.ID: 01279 acc=AL031433
length: 1204 bp unfinished: djl71n11 Contig.ID: 01877
acc=AL031433 Length: 1036 bp Unfinished: djl71n11 Contig.ID:
02399 acc=AL031433 Length: 1401 bp Unfinished: djl71n11
Contig.ID: 00107 acc=AL031433 Length: 1140 bp Unfinished:
djl71n11 Contig.ID: 01160 acc=AL031433 Length: 1235 bp
Unfinished: djl71n11 Contig.ID: 00958 acc=AL031433 Length: 1635
bp Unfinished: djl71n11 Contig.ID: 02950 acc=AL031433 Length:
1064 bp Unfinished: djl71n11 Contig.ID: 02130 acc=AL031433
Length: 1566 bp Unfinished: djl71n11 Contig.ID: 01242
acc=AL031433 Length: 129228 bp Unfinished: djl71n11 Contig.ID:
02096 acc=AL031433 Length: 1269 bp Unfinished: djl71n11
Contig.ID: 02350 acc=AL031433 Length: 1133 bp Unfinished:
djl71n11 Contig.ID: 00120 acc=AL031433 Length: 1026 bp
Unfinished: djl71n11 Contig.ID: 01394 acc=AL031433 Length: 1103
bp Unfinished: djl71n11 Contig.ID: 00468 acc=AL031433 Length:
1142 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
Location/Qualifiers
1..155582

```

BASE COUNT	a	c	g	t	others
ORIGIN	44935	29770	28464	42011	10402

Query Match	84.28;	Score 16;	DB 19;	Length 155582;
Best Local Similarity	100.08;	Pred. No. 2.36e+01;		
Matches	16;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

```

Db 29979 ACAATATTAAAGAAG 29994
      |||
QY 4 ACAATATTAAAGAAG 19

```

RESULT	9				
LOCUS	AC002449	173767 bp	DNA	PRI	20-AUG-1997
DEFINITION	Human PAC clone DJ404K21 from xq23, complete sequence.				
ACCSSION	AC002449				
NID	g2337886				
VERSION	AC002449.1	GI:2337886			
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.				
REFERENCE	1 (bases 1 to 173767)				
AUTHORS	Hinds, K., Tin-Wollam, A., Sutterer, C and Fromick, B.				
TITLE	The sequence of H. sapiens PAC clone DJ404K21				
JOURNAL	Unpublished (1997)				
REFERENCE	2 (bases 1 to 173767)				
AUTHORS	Waterston, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-AUG-1997) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA				
COMMENT	SUBMITTED BY:				

Department of Genetics  
Washington University  
St. Louis MO 63108, USA  
<http://genome.wustl.edu/gsc>  
<mailto:sapiens@watson.wustl.edu>



NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by David Bentley's chromosome X mapping group at the Sanger Centre, Wellcome Trust Genome Campus, Hinxton, UK. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>

#### SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-3 prepared by Pieter de Jong and coworkers at Roswell Park Cancer Institute, using the method described by Ioannou et al., Nature Genetics, 6:84-9 (1994). The library is from one male donor. For further details, see <http://bacpac.med.buffalo.edu/> The clone is available from Genome Systems, Inc. (<http://www.genomesystems.com>).  
VECTOR: pCYPAC2

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is H.DJ790B04. The actual start of this clone is at base position 1 of DJ404K21; actual end is at 173767 of DJ404K21. This clone is part of an unanchored island, orientation is unknown.

This clone contains STS's SWXD757 (NTD:9405440) and SWXD512 (NTD:9995449).

#### FEATURES

##### Source

Location/Qualifiers  
1. 173767  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/clone="DJ404K21"  
/clone\_11b="RPCI-3"  
/map="Xq23"  
3958..3983  
repeat\_region  
/rpt\_family="L1"  
7187..7448  
repeat\_region  
/rpt\_family="MER"  
7450..7832  
repeat\_region  
/rpt\_family="ALU"  
complement(7833..7891)  
repeat\_region  
/rpt\_family="THR"  
7893..8184  
repeat\_region  
/rpt\_family="ALU"  
complement(8198..9670)  
repeat\_region  
/rpt\_family="THR"  
9671..10045  
repeat\_region  
/rpt\_family="ALU"  
10047..10226  
repeat\_region  
/rpt\_family="MER"  
10461..11065  
repeat\_region  
/rpt\_family="L1"  
complement(11071..11286)  
repeat\_region  
/rpt\_family="MER"  
11291..11433  
repeat\_region  
/rpt\_family="L1"  
complement(11434..11593)  
repeat\_region  
/rpt\_family="MER"  
complement(11616..11641)  
repeat\_region  
/rpt\_family="MER"  
complement(114472..114820)  
repeat\_region  
/rpt\_family="L1"

repeat\_region complement(14833..15202)  
/rpt\_family="L1"  
repeat\_region 15777..15808  
/rpt\_family="L1"  
repeat\_region 16030..16436  
/rpt\_family="L1"  
repeat\_region 16625..16719  
/rpt\_family="L1"  
repeat\_region 16860..18448  
/rpt\_family="L1"  
repeat\_region 18450..18738  
/rpt\_family="ALU"  
repeat\_region 18756..22756  
/rpt\_family="L1"  
repeat\_region complement(24561..24819)  
/rpt\_family="ALU"  
repeat\_region 25322..25360  
/rpt\_family="L1"  
repeat\_region 25871..25936  
/rpt\_family="L1"  
repeat\_region 27205..27559  
/rpt\_family="L1"  
repeat\_region 27921..28394  
/rpt\_family="L1"  
repeat\_region 29212..29235  
/rpt\_family="L1"  
repeat\_region complement(30468..30557)  
/rpt\_family="L1"  
misc\_feature 30844..31126  
/note="match to EST R00201 (NTD:9749937) ye71f11.r1"  
misc\_feature 31050..31156  
/note="match to EST R00201 (NTD:9749937) ye71f11.r1"  
repeat\_region 31530..32115  
/rpt\_family="L1"  
repeat\_region complement(34006..34295)  
/rpt\_family="ALU"  
repeat\_region 35839..35865  
/rpt\_family="L1"  
repeat\_region 36463..37291  
/rpt\_family="L1"  
repeat\_region 39129..39159  
/rpt\_family="L1"  
repeat\_region complement(39829..40254)  
/rpt\_family="ALU"  
repeat\_region 40512..41114  
/rpt\_family="L1"  
repeat\_region 41227..41458  
/rpt\_family="L1"  
repeat\_region complement(41461..41750)  
/rpt\_family="ALU"  
repeat\_region 42582..42630  
/rpt\_family="L1"  
repeat\_region 43534..43661  
/rpt\_family="L1"  
repeat\_region 44280..44440  
/rpt\_family="L1"  
repeat\_region 44494..44678  
/rpt\_family="L1"  
repeat\_region complement(45130..45422)  
/rpt\_family="ALU"  
repeat\_region complement(45972..46009)  
/rpt\_family="THR"  
repeat\_region 47644..47668  
/rpt\_family="L1"  
repeat\_region 48498..48788  
/rpt\_family="ALU"  
repeat\_region complement(49482..49770)  
/rpt\_family="ALU"  
repeat\_region complement(51436..51509)  
/rpt\_family="L1"  
repeat\_region complement(51533..51623)  
/rpt\_family="L1"  
repeat\_region complement(52389..52862)



```

repeat_region /rpt_family="LI"
complement(52874, .54009)
repeat_region /rpt_family="LI"
54033, .54244
repeat_region /rpt_family="LI"
complement(54308, .54351)
repeat_region /rpt_family="LI"
54388, .54524
repeat_region /rpt_family="ALU"
54525, .57141
repeat_region /rpt_family="LI"
57168, .58229
repeat_region /rpt_family="LI"
58231, .58520
repeat_region /rpt_family="ALU"
58522, .61300
repeat_region /rpt_family="LI"
61313, .61462
repeat_region /rpt_family="ALU"
complement(61490, .61545)
repeat_region /rpt_family="LI"
complement(61550, .61830)
repeat_region /rpt_family="ALU"
complement(61833, .61952)
repeat_region /rpt_family="LI"
63130, .63422
repeat_region /rpt_family="ALU"

```

Note: remainder of annotations omitted.

```

Query Match      84.2%; Score 16; DB 30; Length 173767;
Best Local Similarity 100.0%; Pred. No. 2.36e+01;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db 102109 TACAAATATTAAAGA 102124

Qy 3 TACAAATATTAAAGA 18

```

RESULT 10
LOCUS HS288M22 182606 bp DNA HTG 24-FEB-1999
DEFINITION Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone
288M22, WORKING DRAFT SEQUENCE.
ACCESSION AL035467
NID 94455492
VERSION AL035467.1 GI:4455492
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 182606)
AUTHORS Sulston, J.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humanyesanger.ac.uk Clone requests: clonerequestsanger.ac.uk
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phase etc. Order of segments is not known: 800 n's separate
segments. Unfinished: dj288M22 Contig_ID: 01639 acc- length:
17052 bp Unfinished: dj288M22 Contig_ID: 02002 acc- length:
47937 bp Unfinished: dj288M22 Contig_ID: 00927 acc- length:
24720 bp Unfinished: dj288M22 Contig_ID: 01525 acc- length:
45863 bp Unfinished: dj288M22 Contig_ID: 01820 acc- length:
18665 bp Unfinished: dj288M22 Contig_ID: 01565 acc- length: 1148
bp Unfinished: dj288M22 Contig_ID: 01835 acc- length: 1210 bp
Unfinished: dj288M22 Contig_ID: 01180 acc- length: 1288 bp
Unfinished: dj288M22 Contig_ID: 01179 acc- length: 1241 bp
Unfinished: dj288M22 Contig_ID: 00874 acc- length: 1220 bp

```

```

Unfinished: dj288M22 Contig_ID: 00889 acc- length: 1366 bp
Unfinished: dj288M22 Contig_ID: 02288 acc- length: 1305 bp
Unfinished: dj288M22 Contig_ID: 02707 acc- length: 1305 bp
Unfinished: dj288M22 Contig_ID: 01546 acc- length: 1113 bp
Unfinished: dj288M22 Contig_ID: 00528 acc- length: 1427 bp
Unfinished: dj288M22 Contig_ID: 01676 acc- length: 1656 bp
Unfinished: dj288M22 Contig_ID: 02676 acc- length: 1250 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
Location/Qualifiers
source 1..182606
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="288M22"

```

```

BASE COUNT 55348 a 31397 c 30962 g 52083 t 12816 others
ORIGIN

```

```

Query Match      84.2%; Score 16; DB 19; Length 182606;
Best Local Similarity 100.0%; Pred. No. 2.36e+01;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db 69195 TCTTTAATATTGTAA 69210

Cp 17 TCTTTAATATTGTAA 2

```

RESULT 11
LOCUS G14874 250 bp DNA STS 04-JAN-1996
DEFINITION human STS SHGC-13859.
ACCESSION G14874
NID G1131637
VERSION G14874.1 GI:1131637
KEYWORDS STS sequence; primer; sequence tagged site.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata;
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 250)
AUTHORS Myers, R.M.
JOURNAL Unpublished (1995)
COMMENT

```

Contact: Richard M. Myers  
Stanford Human Genome Center (SHGC)  
Stanford University School of Medicine  
Department of Genetics, M-344, Stanford, CA 94305, USA  
Tel: 4157259687  
Fax: 4157259689  
Email: myerseshgc.stanford.edu

Primer A: TACGCAAACTCATGGAAGC  
Primer B: TCCGTCACATCATATGCTTT  
STS size: 136  
PCR profile:

Initial incubation: 94 degrees C for 90 seconds  
Denaturation: 94 degrees C for 15 seconds  
Annealing: 62 degrees C for 23 seconds  
Polymerization: 72 degrees C for 30 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9600  
Protocol:

Template: 25 ng  
each 1 uM  
Primer: each 200 uM  
dNTPs: each 200 uM  
Tag polymerase: 0.05 units/ul  
Total Vol: 10 ul

Buffer:



MgCl2: 2.5 mM  
KCl: 50 mM  
Tris-HCl: 20 mM  
pH: 8.3

Prepared with primer pairs derived from T55202--Merck/Unilest.  
Location/Qualifiers

FEATURES  
Source  
1. .250  
/organism="Homo sapiens"  
STS  
primer\_bind 114. .249  
primer\_bind 114. .135  
complement(227. .249)  
BASE COUNT 109 a 38 c 40 g 62 t 1 others  
ORIGIN

Query Match 78.9%; Score 15; DB 34; Length 250;  
Best Local Similarity 100.0%; Pred. No. 8.82e+01;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 79 ACAATATTAAAGAA 93  
|||||  
OY 4 ACAATATTAAAGAA 18

RESULT 12  
LOCUS AF010534 429 bp DNA INV 16-JUL-1998  
DEFINITION Plasmodium falciparum microsatellite TA22 sequence.  
ACCESSION AF010534  
NID 92253526  
VERSION AF010534.1 GI:2253526  
KEYWORDS  
SOURCE  
ORGANISM  
Malaria parasite P. falciparum.  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE  
1 (bases 1 to 429)  
Su, X. z. and Wellem, T. E.  
TITLE  
Toward a high-resolution Plasmodium falciparum linkage map:  
polymorphic markers from hundreds of simple sequence repeats  
JOURNAL  
Genomics 33 (3), 430-444 (1996)  
MEDLINE  
96299638  
REFERENCE  
2 (bases 1 to 429)  
Su, X. z. and Wellem, T. E.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (27-JUN-1997) NIAID-Laboratory of Parasitic Diseases,  
NIH, Bethesda, MD 20892, USA  
FEATURES  
Location/Qualifiers  
1. .429  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:5833"  
/chromosome="14"  
repeat\_region 1. .429  
/note="microsatellite TA22"  
BASE COUNT 199 a 21 c 30 g 179 t  
ORIGIN

Query Match 78.9%; Score 15; DB 22; Length 429;  
Best Local Similarity 100.0%; Pred. No. 8.82e+01;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 15 TTACAATATTAAAG 29  
|||||  
OY 2 TTACAATATTAAAG 16

RESULT 13  
LOCUS AF127454 921 bp DNA PLN 15-MAR-1999  
DEFINITION Peumus boldus rp16 gene, chloroplast gene encoding chloroplast  
protein, intron sequence.  
ACCESSION AF127454  
NID 94416513  
VERSION AF127454.1 GI:4416513  
KEYWORDS  
SOURCE  
Peumus boldus.  
ORGANISM  
Chloroplast Peumus boldus

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Laurales;  
Monimiacae; Peumus.  
1 (bases 1 to 921)  
Renner, S. S.  
TITLE  
Circumscription and phylogeny of Laurales: evidence from molecular  
and morphological data  
JOURNAL  
2 (bases 1 to 921)  
REFERENCE  
Renner, S. S.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (11-FEB-1999) Biology, Univ. of Missouri-St. Louis, 8001  
Natural Bridge Rd., St. Louis, MO 63121, USA  
FEATURES  
Location/Qualifiers  
1. .921  
/organism="Peumus boldus"  
/chloroplast  
/db\_xref="taxon:63812"  
1. .921  
/gene="rp16"  
1. .921  
/gene="rp16"  
intron  
BASE COUNT 255 a 192 c 139 g 334 t 1 others  
ORIGIN

Query Match 78.9%; Score 15; DB 28; Length 921;  
Best Local Similarity 100.0%; Pred. No. 8.82e+01;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 541 TCTTAATATTGTA 555  
|||||  
Cp 17 TCTTAATATTGTA 3

RESULT 14  
LOCUS HUMPS1A10 1003 bp DNA PRI 21-APR-1997  
DEFINITION Homo sapiens presentinlin (PSI.S182) gene, exon 12.  
ACCESSION L76527  
NID 91479969  
VERSION L76527.1 GI:1479969  
KEYWORDS  
presentinlin.  
10 of 11  
SEGMENT  
SOURCE  
Homo sapiens (clone: PAC clone RPCI 54D12) adult DNA.  
ORGANISM  
Homo sapiens  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
1 (bases 1 to 1003)  
Sherrington, R., Rogaev, E. I., Liang, Y., Rogaeva, E. A., Levesque, G.,  
Ikeda, M., Chi, H., Lin, C., Li, G., Holman, K. et al.  
TITLE  
Cloning of a gene bearing missense mutations in early-onset  
familial Alzheimer's disease  
JOURNAL  
Nature 375 (6534), 754-760 (1995)  
MEDLINE  
95319502  
REFERENCE  
2 (bases 1 to 1003)  
Rogaev, E. I., Sherrington, R., Rogaeva, E. I., Levesque, G., Liang, Y.,  
Ikeda, M., Tsuda, T., Rogaeva, E., Chi, H., Holman, K., Rommens, J. M. and  
St George-Hyslop, P. H.  
TITLE  
Genomic structure, promoter sequence, and alternative splicing of  
the presentinlin-1 gene associated with early onset Alzheimer's  
Disease  
JOURNAL  
Unpublished, (1996)  
Location/Qualifiers  
1. .1003  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="PAC clone RPCI 54D12"  
/dev\_stage="adult"  
/germline  
/map="14q24.3"  
order(L76526.1:625. .>1090.<1. .430)  
/number=11  
431. .549  
/gene="PSI"

JOURNAL  
FEATURES  
Source  
1. .1003  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="PAC clone RPCI 54D12"  
/dev\_stage="adult"  
/germline  
/map="14q24.3"  
order(L76526.1:625. .>1090.<1. .430)  
/number=11  
431. .549  
/gene="PSI"



/note="corresponds to bp 1378-1497 of L42110; putative"  
/citation=[2]  
/number=12

BASE COUNT 292 a 188 c 193 g 313 t 17 others

ORIGIN

Query Match 78.9% Score 15; DB 30; Length 1003;  
Best Local Similarity 100.0%; Pred. No. 8.82e+01;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 403 TTTAATTTTGTAC 417  
|||||

Cp 15 TTTAATTTTGTAC 1

RESULT 15  
LOCUS STINVERTN 1911 bp DNA PLN 28-NOV-1996  
DEFINITION S.tuberosum Invertase gene (1911bp).  
ACCESSION X95820  
NID 91695191  
VERSION X95820.1 GI:1695191  
KEYWORDS beta-fructofuranosidase; Invertase.  
SOURCE potato.  
ORGANISM Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Asteridae; Solananae; Solanales; Solanaceae; Solanum; Potatoe;  
section Petota.  
1 (bases 1 to 1911)  
Bounnay,A.S., Hedley,P.E., Maddison,A., Waugh,R. and Machray,G.C.  
Exon skipping induced by cold stress in a potato Invertase gene  
transcript  
Nucleic Acids Res. 24 (12), 2347-2351 (1996)  
2 (bases 1 to 1911)  
Maddison,A.L.  
Direct Submission  
Submitted (23-FEB-1996) A.L. Maddison, Scottish Crop Research  
Institute, Cell and Molecular Genetics, Invergowrie, Dundee, DD2  
5DA, UK  
revised by [3]  
3 (bases 1 to 1911)  
Maddison,A.L.  
Direct Submission  
Submitted (28-NOV-1996) A.L. Maddison, Scottish Crop Research  
Institute, Cell and Molecular Genetics, Invergowrie, Dundee, DD2  
5DA, UK  
On Nov 29, 1996 this sequence version replaced gi:1694912.  
Related sequence Z21486.  
Location/Qualifiers  
1. 1911  
/organism="Solanum tuberosum"  
/cultivar="Cara"  
/db\_xref="taxon:4113"  
<1. 20  
/number=1  
join(<1. 20,1762. 1770,1893. >1911)  
/ec\_number="3.2.1.26"  
/note="beta-fructofuranosidase"  
/codon\_start=2  
/product="invertase"  
/protein\_id="CAA65086.1"  
/db\_xref="PID:e283687"  
/db\_xref="PID:g1695192"  
/db\_xref="GI:1695192"  
/db\_xref="SPTREMBL:Q43172"  
/translation="PRKWINDPNAPMYNG"  
21. 1761  
/number=1  
1762. 1770  
/number=2  
1771. 1892  
/number=2

exon  
1893. >1911  
/number=3

BASE COUNT 615 a 266 c 274 g 756 t

ORIGIN

Query Match 78.9% Score 15; DB 27; Length 1911;  
Best Local Similarity 100.0%; Pred. No. 8.82e+01;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1274 TACAAATATTAAAGA 1288  
|||||

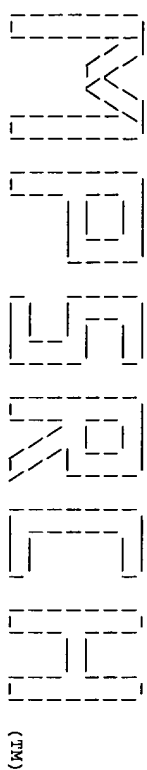
Qy 3 TACAAATATTAAAGA 17

Search completed: Sat Nov 27 15:19:43 1999  
Job time : 253 secs.



**THIS PAGE BLANK (USPTO)**





Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPearch\_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 15:25:27 1999; MasPar time 0.62 Seconds  
90.461 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-6  
Description: (1-19) from US09103287.seq  
Perfect Score: 19  
N.A. sequence: 1 GTTACAAATATTAAAGAG 19  
Comp: CAATGTTATTAATTTCTTC

Scoring table: TABLE jmetric  
Gap 60

Mmatch STD : Dbase 0; Query 0

Searched: 1052 seqs, 1486975 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: HIV-NAB  
1:H\_PRI 2:H\_UNA 3:H\_VIR

Statistics: Mean 7.179; Variance 1.586; scale 4.525

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	12	63.2	1396	3	HIVNH52	Human immunodeficiency	1.18e+00
2	12	63.2	3983	3	HIVSF2B13	Human immunodeficiency	1.18e+00
3	12	63.2	4273	3	HIVSC	Human immunodeficiency	1.18e+00
4	12	63.2	4684	3	HIVP896	Human immunodeficiency	1.18e+00
5	12	63.2	9190	3	HIVORY	Human immunodeficiency	1.18e+00
6	12	63.2	9540	3	HIVRCST	Human immunodeficiency	1.18e+00
7	12	63.2	9737	3	HIVSF2	Human immunodeficiency	1.18e+00
8	11	57.9	1773	3	SIVAGMT49	Simian immunodeficiency	7.89e+00
9	11	57.9	2583	3	HIVZCAM4E	Human immunodeficiency	7.89e+00
10	11	57.9	3120	3	HIV14575	Human immunodeficiency	7.89e+00
11	11	57.9	9170	3	SIVAGMTYO	Simian (African green	7.89e+00
12	11	57.9	10036	3	SIVSABIC	Simian immunodeficiency	7.89e+00
13	10	52.6	1119	3	SIVRAN40E	Simian immunodeficiency	4.19e+01
14	10	52.6	1478	1	HUMMS1	Human mRNA for the MS	4.19e+01
15	10	52.6	1512	3	HIVZD808	Human immunodeficiency	4.19e+01
16	10	52.6	1521	3	HIVZD1024	Human immunodeficiency	4.19e+01
17	10	52.6	1524	3	HIVZD766	Human immunodeficiency	4.19e+01
18	10	52.6	1525	3	HIVZD868	Human immunodeficiency	4.19e+01
19	10	52.6	1600	3	HIVCAR4D20	Human immunodeficiency	4.19e+01
20	10	52.6	2550	3	HIVZCAM5E	Human immunodeficiency	4.19e+01

21	10	52.6	2556	3	HIVZCB122E	Human immunodeficiency	4.19e+01
22	10	52.6	2577	3	HIVZCAM1E	Human immunodeficiency	4.19e+01
23	10	52.6	2580	3	HIVZCB123E	Human immunodeficiency	4.19e+01
24	10	52.6	2583	3	HIVZCAM3E	Human immunodeficiency	4.19e+01
25	10	52.6	2589	3	HIVZCAM6E	Human immunodeficiency	4.19e+01
26	10	52.6	2589	3	HIVZAL1	Human immunodeficiency	4.19e+01
27	10	52.6	2601	3	HIVZCB124E	Human immunodeficiency	4.19e+01
28	10	52.6	3201	3	HIVBAL2	Human immunodeficiency	4.19e+01
29	10	52.6	3600	3	HIVBRVA	Human immunodeficiency	4.19e+01
30	10	52.6	3808	3	HIVBAL1	Human immunodeficiency	4.19e+01
31	10	52.6	8344	3	EIAV	Bovine infection and	4.19e+01
32	10	52.6	8391	3	BIV106	Bovine immunodeficiency	4.19e+01
33	10	52.6	8482	3	BIV127	Bovine immunodeficiency	4.19e+01
34	10	52.6	9431	3	HIVZNIH2	Human immunodeficiency	4.19e+01
35	10	52.6	9468	3	FIVPFR	Feline immunodeficiency	4.19e+01
36	10	52.6	9474	3	FIV14	Feline immunodeficiency	4.19e+01
37	10	52.6	9480	2	HIVZGH1	Human immunodeficiency	4.19e+01
38	10	52.6	9485	3	FIVZ1	Feline immunodeficiency	4.19e+01
39	10	52.6	9591	3	HIVCAM1	Human immunodeficiency	4.19e+01
40	10	52.6	9671	3	HIVZROD	Human immunodeficiency	4.19e+01
41	10	52.6	9672	2	HIVZST	Human immunodeficiency	4.19e+01
42	10	52.6	9754	3	HIVANT70	Human immunodeficiency	4.19e+01
43	10	52.6	10277	3	SIVM251	Simian (macaque) immu	4.19e+01
44	10	52.6	11443	3	HIVZCAM2	Human immunodeficiency	4.19e+01
45	10	52.6	13068	3	SIVM239	Simian immunodeficiency	4.19e+01

# ALIGNMENTS

RESULT	1	HIVNH52	1396 bp ss-DNA	VRL	22-SEP-1992
LOCUS					
DEFINITION					
Human immunodeficiency virus type 1, clone NH52, vif, vpr, vpu, tat regions.					
ACCESSION					
M00000					
SEGMENT					
2 of 3					
SOURCE					
ORGANISM					
Human immunodeficiency virus type 1 (HIV-1), clone NH52.					
Viridae; ss-RNA enveloped viruses; Positive strand RNA virus; Retroviridae; Lentivirinae.					
REFERENCE					
1 (bases 1 to 1396)					
Weidt,G., Falstet,A.C., Sauermann,U., Lueke,W., Jentsch,K.D. and Hummann,G.					
Characterization of a new German HIV-1 isolate: HIV-INHS					
JOURNAL					
Unpublished (1992)					
STANDARD					
Full staff entry					
COMMENT					
Kindly provided prior to publication in computer-readable form by Arne Falstet, Deutsches Primatenzentrum (DPZ), Abteilung fur Virologie und Immunologie, Kellnerweg 4, W-3400 Goettingen, Germany.					
FEATURES					
pept	71	295			
from					
to/span					
description					
vif polypeptide (AA at 71; premature stop at position 293)					
pept.ps	71	649			
pept	589	879			
pept	1091	1159			
pept.ps	1091	1336			
pept	860	1074			
pept	999	1074			
pept	502	230			
BASE COUNT					
ORIGIN					
Query Match					
Best Local Similarity 100.0%;					
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
Db	1100	ttacaataatta 1111			
QY	2	TTCAAAATATTA 13			
RESULT	2	HIVSF2B13	3983 bp ss-RNA	VRL	18-SEP-1992
LOCUS					
DEFINITION					
Human immunodeficiency virus type 1, isolate SF13, tat, rev, vpu,					



```

ACCESSION      M00000
SOURCE          Human immunodeficiency virus type 1 (HIV-1), SF13, isolated from
                the same patient that produced SF2; cocultured with PBMCs and
                cloned into lambda phage; subclone pUC19 DNA.
ORGANISM        Human immunodeficiency virus type 1
                Viridae: ss-RNA enveloped viruses; Positive strand RNA virus:
                Retroviridae; Lentivirinae.
REFERENCE       1 (bases 1 to 3983)
AUTHORS        Cheng-Mayer,C., Shioda,T. and Levy,J.A.
TITLE          Host range, replicative, and cytopathic properties of human
                immunodeficiency virus type 1 are determined by very few amino acid
                changes in tat and gp120
JOURNAL        J. Virol. 65, 6930-6941 (1991)
STANDARD       Full staff-entry
COMMENT        The SF13 viral isolate was recovered from the same individual that
                produced SF2 (see <HIVSF2>), but five months later when the
                individual developed KS and PCP. The increased pathogenicity and
                wider host range of SF13 is studied in depth in [1]: two aa changes
                in tat and 22 aa changes in env, going from SF2 to SF13, are of
                special significance.
FEATURES        from to/span description
                91      305      tat protein, exon 2, (first expressed exon)
                2630     2720     tat protein, exon 3, (AA at 2630)
                230      305      rev protein, exon 2, (first expressed exon)
                2630     2904     rev protein, exon 3, (AA at 2631)
                322      567      vpu protein
                485      3046     envelope polypeptide
                3048     3680     nef protein
                3349     3982     3' LTR
                1338 a    716 c    980 g    949 t
BASE COUNT      1338 a    716 c    980 g    949 t
ORIGIN          Query Match
                Best Local Similarity 100.0%; Score 12; DB 3; Length 3983;
                Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
                2 TTACAAATATTA 13
Db 331 ttacaatat 342
|||||
Oy 2 TTACAAATATTA 13
RESULT 3
LOCUS          HIVSC 4273 bp ss-RNA VRL 01-MAR-1999
DEFINITION     Human immunodeficiency virus type 1, isolate SC (3' end of
                genome).
ACCESSION      M17450
SOURCE          Human immunodeficiency virus type 1 (HIV-1), isolate SC, proviral
                DNA.
REFERENCE       1 (bases 1 to 4273)
AUTHORS        Gurgo,C., Guo,H.-G., Franchini,G., Aldovini,A., Collalti,E.,
                Farrell,K., Wong-Staal,F., Gallo,R.C. and Reltz,M.S.Jr.
TITLE          Envelope sequences of two new United States HIV-1 isolates
JOURNAL        Virology 164, 531-536 (1988)
STANDARD       full staff-review
COMMENT        Kindly made available in computer readable form by Marv Reltz,
                N.C.I., Bethesda, MD 20892 U.S.A. This isolate was taken from a
                California AIDS patient in 1984. There is an in-frame stop at
                position 3212 of the envelope coding sequence; the nef cds is
                uncertain beyond position 4049. A in-frame stop codon, 'taa,' is
                found at positions 4224-4226 in the nef cds..
                location/Qualifiers
FEATURES        prim_transcript <1..>4273
                /note="genomic mRNA"
                prim_transcript <1..>4273
                /note="tat, rev, nef subgenomic mRNA"
                intron <1..487
                /note="tat, rev, nef subgenomic mRNA intron 1"
                CDS <1..330
                /note="vif protein (partial; AA at 1)"
                /codon_start=1
                CDS join(541..755,3089..3179)

```

```

CDS            /note="tat protein"
                /codon_start=541
                join(680..755,3089..3363)
                /note="rev protein"
                /codon_start=680
                270..560
                /note="vpr protein"
                /codon_start=270
                541..755
                /note="tat protein, exon 2 (first expressed exon)"
                680..755
                /note="rev protein, exon 2 (first expressed exon)"
                756..3088
                /note="tat cds intron 2"
                756..3088
                /note="rev cds intron 2"
                756..3088
                /note="tat, rev, nef subgenomic mRNA intron 2"
                772..864
                /note="vpu protein (premature termination)"
                /codon_start=772
                935..3505
                /pseudo
                /note="envelope polypeptide (premature stop at 3212)"
                /codon_start=935
                3089..3179
                /note="tat protein, exon 3 (AA at 3090)"
                3089..3363
                /note="rev protein, exon 3 (AA at 3091)"
                3212..3214
                /note="premature stop (tag) in env cds"
                /note="premature stop (tag) in env cds"
                3507..4226
                /note="nef protein"
                /codon_start=3507
                3796..>4273
                /note="3' LTR"
                4249..>4273
                repeat_region
                /note="R repeat 3' copy"
                760 c    1053 g    1013 t
BASE COUNT      1447 a    760 c    1053 g    1013 t
ORIGIN          Query Match
                Best Local Similarity 100.0%; Score 12; DB 3; Length 4273;
                Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
                2 TTACAAATATTA 13
Db 781 ttacaatat 792
|||||
Oy 2 TTACAAATATTA 13
RESULT 4
LOCUS          HIVP896 4684 bp ss-RNA VRL 24-APR-1993
DEFINITION     Human immunodeficiency virus type 1 proviral DNA encoding the vif,
                vpr, vpu, env, tat, rev and nef genes, and 3' LTR, clone p89.6
                from strain 89.6.
ACCESSION      M96155
SOURCE          Human immunodeficiency virus type 1 (HIV-1), clone p89.6,
                proviral DNA.
ORGANISM        Human immunodeficiency virus type 1
                Viridae: ss-RNA enveloped viruses; Positive strand RNA virus:
                Retroviridae; Lentivirinae.
REFERENCE       1 (bases 1 to 4684)
AUTHORS        Collman,R., Balliet,J.W., Gregory,S.A., Friedman,H., Kolson,D.L.,
                Nathanson,N. and Srinivasan,A.
TITLE          An infectious molecular clone of an unusual macrophage-tropic and
                highly cytopathic strain of HIV-1
JOURNAL        J. Virol. 66, 7517-7521 (1992)
STANDARD       full staff-review
COMMENT        The authors point out that this sequence is unusual because it is
                both macrophage-tropic and extremely cytopathic in lymphocytes:
                they contend that the virus is unusual in so far as it is
                the first well-characterized infectious molecularly cloned
                macrophage-tropic HIV-1 strain derived from peripheral blood.

```



The HIV896 env sequence clusters with HIV-1 subtype B env sequences.

```
FEATURES
CDS
    Location/Qualifiers
        join(800..1014,3339..3429)
        /gene="tat"
        /codon_start=-1
        /translation="MEPVNPSPLEPMKHPGSSQPKTACTNCYCKKCCFHCACGCTTGKLG
        ISYGRKRQRQRRPQDSQTHQVSLSKSPSSQPRGPTPKDKKKVRETFETDPVH"
        join(939..1014,3339..3613)
        /gene="rev"
        /codon_start=-1
        /translation="MAGRSQSDSDLDLKVRLIKFIYQSNPPSLSTQARNRNR
        MWRQRQISISIRILGFLGRFEEFVPLPLPLEKLTLIDCNDGCTSGTQGVSPQI
        LVESPAILEPRTKE"
        11..589
        /gene="vif"
        /codon_start=-1
        /translation="MENRQVMTVMQVDMKRTITKSLYKHHMTISGKAKGSYRHH
        ESTNPRISSVHPIGDADLVTTTGLTGERDWMVGGSISEMKRRYSQVDPGL
        ADRLHLIYFDCSDSAIKRSILGHIVSPSCYOAGHNKVGSLQYLALALATTPRRIK
        PEPSPYTKLTEDRMNPKQTKGHSHTMTGH"
        529..819
        /gene="vpr"
        /codon_start=-1
        /translation="MEQAPEDQGPQREPYNDWTLELLEELKNEAVHPRIMLSIQG
        HIYETIGDTWGEALIRILOQLLFIHFIRIGCHRSRIGIIOHRTENGASKS"
        1015..3338
        /gene="tat"
        1015..3338
        /gene="rev"
        1031..1273
        /gene="vpu"
        /codon_start=-1
        /translation="MLSQILAIVALVAIIAYVWVIVFIEYKILNRKIDRLID
        RIERRERDSGNSGQDEELALERHLPWVDL"
        1194..3755
        /product="envelope polypeptide"
        /gene="env"
        /codon_start=-1
        /translation="MRVKEIRKNMOMLRGGILLGLMIMCSAKERTWTIYGVPM
        REATITLPCASAKADTVENHVMATHACVPDPNQEYVLGVNTENPMKNNMYDQ
        MHEDITSLMDESKPCVKLPLCVTLNLTNTNTNTPTSSWGMKGEIKNSF
        YTTISIRNKVKKEALFNLDVPLENTNTNRYNRNCTSVITQCPKPSQDPIH
        YCPVPAFAMLKCNKNTFNKSGPCTNCTVCTGILPVSSTOLLGSLAEDIVRS
        ENFTDAKTIIVOLNESVYINCTRPNNNTRRISIGRPAFARNIIGDIOAHONI
        SAKNNNTLOQIVIKLREKFNKTIAFNOSGSDPEIVHSGGGEFFYCTAOFIN
        SNWYVTCNGTGEENDITLQCRKQIIMMOKVGMAMAPPTGQIRRSNTGILL
        TPDGNGSTETETETIRPPGGDMNMRSELITKVRIPPIGVAIPRARRKRVOREKR
        AVGIGAVLGFAGAGSTMGARSVLTQARLLSGIVQOQNNLALIAQOQHMQLT
        VWGIGQARVALAERYLDQOLMGIMGSGKLICTTSPMVMSSNKSQVDIMNMT
        NMEEREINDNTYDIYDLKESQTOEKNEKELLEDKASLMNMDINLMYETIE
        IMVGGILGLRIYFAVLSIVNVRGQSPSLFOTLIPASRGDPGCTEEGGERD
        RSGPLVNGFLAFWVDLRNLCFLYHLLNLLIYRIYELLGRKRWELKTKWNLQ
        IWSQELKNSAVSLNATAIYAVGSDTVIKIVQACRAIRNIPTRIRQGLERALL"
        1194..2717
        /product="glycoprotein 120"
        /gene="env"
        /codon_start=-1
        /product="glycoprotein 120"
        2718..3752
        /product="glycoprotein 41"
        /gene="env"
        /codon_start=-1
        3757..4371
        /gene="nef"
        /codon_start=-1
        /translation="MGGKSKRAESWQITIREMRAPAEPAADGCVGAVSNDLARHG
        ATTSSTNNADIAWLEAQEGEVEGPRVQVPLRPPTYAAADLSFLFKEGKDELY
        HSQKQDILDLWVYHTQFFPDMDNTPPGGIRYPLTFGMCNKLVPVEBDEGENNED
        NSLHPANQHGVEDSERQYLVWRFDSRLAFHHVARELHREYFKN"
        4049..4684
        4049..4502
        LTR
        repeat_region
        /rpt_type=other
        /note="03 region"
        /note="03 region"
```

```
repeat_region 4503..4599
    /rpt_type=other
    /note="R region"
repeat_region 4600..4684
    /rpt_type=other
    /note="05 region"
source 1..4684
    /organism="Human immunodeficiency virus type 1"
    /note="Human peripheral blood mononuclear cells."
    /provisional
    /sequenced_mol="DNA"
    /issue="lib="subgenomic lambda phage library of HMW"
BASE COUNT 1579 a 849 c 1145 g 1111 t
ORIGIN 11 bp downstream of vif cds start
Query Match 63.28; Score 12; DB 3; Length 4684;
Best Local Similarity 100.08; Pred. No. 1.18e+00;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1040 ttacaataatata 1051
QY 2 TTACAATAATATA 13
RESULT 5
LOCUS HIVOYI 9190 bp ss-RNA VRL 15-OCT-1991
DEFINITION Human immunodeficiency virus type 1, isolate OYI, complete genome.
ACCESSION M26727
SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate OYI proviral
DNA (also designated gab397).
REFERENCE 1 (bases 1 to 9190)
AUTHORS Wain-Hobson, S.
TITLE Unpublished
JOURNAL AIDS Res. Hum. Retroviruses (1990) In press
STANDARD
COMMENT
Kindly provided in computer-readable form by Simon Wain-Hobson,
Pasteur Institute, Paris.
```

OYI, designated elsewhere as isolate 397, was taken from a healthy Gabonese presenting an atypical Western Bloc. Tree analysis (see Part III) reveals that the sequence is closely related to the North American isolate SF2 and the European virus HAM/across the genome). This is the first report of a virus from Africa that clusters with North American rather than Zairean viruses: OYI and SF2 differ by approximately 7% in envelope. The single C->S substitution at residue 22 of the OYI tat protein renders it inactive, but may not account for the avirulence of the virus. The pattern of env termination and nef start is not what is typically encountered. A possible mutation from a 'g' to an 'a' at position 5606 produced the triplet 'ata' instead of the usually conserved start codon for the vpr coding region.

Sibling sequences for OYI(397) are available (see 1989:I-A-181).

```
FEATURES
exon
    Location/Qualifiers
        <5373..5587
        /note="tat protein, exon 2 (first expressed exon)"
        <5512..5587
        /note="rev protein, exon 2 (first expressed exon)"
        <1628..4639
        /note="pol polypeptide"
        /codon_start=-1
        /translation="FERRDIAFPQGRKAREFSQTRAMPSTRRLRVNGGRNNSPFA
        GADRGQTSFNSPQTTIMQRPVYTKIGQKLEALLDGADTVLEENLPERKRPKI
        IGIGGFTKRVYDQILIEIGKRAIGVAVGPTVNIIGRLLQLQCTLNFPISPI
        ETVPVAKLRGMDGPRVKQMPLEEKIKVLEICTMEKRGKISKVGPENPYNTPVFAI
        KKKDSTKRPKMDVDEFLNRKRDQDEWEOVLGIFHPAGLKKSIVLVGDVAVFSPALD
        KDFRYTFTTIPSINNEPGRIOYOVNVLPQGGKSPALFOSSMTLLEDPFRONDDIV
        IYQVMDLVYGSDELIGORHRIEELRHLRWGPTTDRKQKRPPLAMPXGTEHPD
        KWTVPNIMPENQSTVNDIQVLVKLMASQIYAGIVYKNCILKRLTKALTEVITPL
        TEAELEAENREILKEPVHGYVYPSDVLVAELQKQGGQWTVQIYQEPENKATGK
        YARMGAHTNDVKQLEAVOKITQESIIVIGWTPAFKLPQKQETWEAMWTEKQATWI
        PEMEFVNTPELVKWTQLEKDPVIGAEFFYDGAANRETKLGKAGYVDRGQKQVSL
```



```

IDTNTOKTELQAIHLALODSGLEVINIYDSDQYALGCIQAQDPKSSSELSVQIIIEOLIK
KEKYLAWPAHKGIGNEQVYKLVASAGIRKYLFDGIDKAEHEKHSNRAASD
ENLPVVAKEIYASCDKCOLKCEAMHQVDCSPGJQIDCTLBEKILIVAVHVSQY
IEAEVYPAETGOETAYFYLKLAGRWPKTIHNDNSNFTSTVYKRAKMAWJKEQFGI
PYNPOSQGVESMNNELKIIQGVDAQHEHTATVOMAVFINFRKGGIGYSGER
IVDITATQTEKLOKOITKIONFRVYVYDSEPELMKPAKLKMGEGAVIVQDMSDI
KVPFRKAKIINDIGROMAGDDCVASRODED"
Join(5373..5587,7918..8008)
/note="tat protein"
/codon_start=1
/translation="MEVPDRLPEMKHPGSOPTASNNCYKRCCLHCOVCFKGLG
ISYGRKRORRORAPODSKTHOVSLSKQASGPRDDPGPKSKKVERETETDED"
Join(5512..5587,7918..8192)
/note="rev protein"
/codon_start=1
/translation="MAGRSGDSDELLKTVRLIKFYQSNNPPEPNEGTRQARRRRR
WREORJRKISGWLSTYLGSRAPVPLQPLPRLRLNLDSEDCGTSGTGVSPET
LVESPAVLEPCKE"
prim_transcript 1..9190
/note="genomic mRNA"
182..199
/note="primer (Lys-tRNA) binding site"
336..1835
/note="gag polyprotein"
/codon_start=1
/translation="MGARASVLSGGLDKWEMKIRLPQGGKKRYOLKHIWASRELERF
KINEPLETESECRQILGOLPILKSGSEISLNTVATLYICVQKIEPDKTALD
KIEONKSKKAAQOTADTNGSSQVSNYPVOMLOCOMVHVCISQKIEPDKTALD
EKAFSEVIMPSALAGATPODMLTNTGVSQVAMOMKETINEAASAKKIVYR
HAGPAPQOMREPESDIACTSTIOEQGMNTNPPIPVGETIYKRIITLGNKIVRYM
ISPIIDIRQKPEFDYDRFYKTLNABASQDVKNMWTILVONANDCKTIL
KALGPAALKEEMTAICQGVGGHARVLAEMASQVNSTVYOMQGNKNDKTYKCF
NCGKGHLAKNCRAPRKCKGCMKCGREGHOMKDCETROANFLCKIMPISKRGNGNLON
RPEPAPPAESGEGEETTPQKQEPIDKGLYPLTSLRSLGNDPSSQ"
4584..5162
/note="vif protein"
/codon_start=1
/translation="MENRMQVNIWQVDRMIRITWKSLSVKNHMYSKAKGNYRHHY
ESTHRISEVPIPLGDATLVYTYWGLTGEREHLGOGASIEWRKKRYSTOVDPGL
ADOLIHYYFDFCESAIRNALIGNIVPRCEYRPGAHNKVGLQYLAALAIKPKKIK
PPLPSEVTKLEDRMNMKPKQTKGHRGSHMNGH"
5102..5392
/note="vpr protein"
/codon_start=1
/translation="MEQAPEDQGPQREPYNENTLELSEAVRHPRIWLHSLQ
YIYEVGDTWEVEAIIIRILOLLFIRIGCGHSRIGITRORRANGASRS"
5767..8334
/note="env polyprotein"
/codon_start=1
/translation="WTANGTRKNYORLNRWGTMLGMLMISCAENLWVYVGVAV
KEATTLFCASDARAYATEVHVWATACVPTDPQDEVLGNVTEINDMKNMNAEQ
MOEDIIISLMDOSLAKPCVRLTCLVTLCTDVTTSLSRNATNTSSMEMEKELK
NCSEFTTISIRDKMOEYALFKYLDVLPIDKIDTFEFLIHCTSTIIOACPISEPI
PMHCTPGAPGALIKNDKKNGTGPCTNVSTYOCNGHGTIKPVYSTOLLNLSIAEVEI
IRSSFTNNAKIIIVOLKSVYINCTPRNNTRNISTGCPRAFTTIOITGIDTQAH
CNSRATVEKTELEQIATLRKQENKTIADFSSGDEIYHSHSGCEFFYCTSQ
LENSTWMDTTRANSFEVITLPCRIOIVNMQEVGKMAVPIISGQJRSKSTIGL
LTPDQKMTNGIEIFRAGSGMDNRMSLEYKYVVEIEPGLVPTKARVYORER
RAVGLMGALGELFAGSGMTGARSMTLVAROLLSGIVQOONNLLRAIEVOOHLQ
LTVNGIKOLOARVLAVERIKYKIDDOOLGIMGCSKRICTTYPWMAWSNKSINEMW
MTWQWEREIDNYTHLYITLIEESONOEKNOELLELDKNAAGMSNFTSNMWTIR
IFIITVGLVGLRIVFAVLSIVNRQGYSPISFTGRLPTQGPDPGIEEGGERD
RDRSGRLVDGFLALIMDDURSLCLESYHRLDLIIVARIVELLGRGMEVLYKWMNL
LOYWQSELKNSYLSLNTATAIYAGCTDYRVEIYQVRAVAFNLPRIRIOGELRALT"
7918..8008
/note="tat protein, exon 3 (AA at 7919)"
7918..8192
/note="rev protein, exon 3 (AA at 7920)"
8336..8971
/note="nef protein"
/codon_start=1
/translation="MGKSKSCSMKGPFTIRERKRAILOPEPAABGVGAASRLK
HGALTSNTATNADCANLEAOEDEEVGPVRPOVPLRPMYIKGALDLSHLKEKGL

```

```

EGLYSQRQDILDLMWYHIGCYPPDMQNTPGGIRYPLCFGKCEKLVMPDPQVEE
ANEGNNLSLHPISLHGDDPEKEVLWKRFLAFRRMAREVHPREYKDC"
misc_signal 9166..9171
/note="poly-A signal"
BASE COUNT 3304 a 1630 c 2202 g 2054 t
ORIGIN

```

```

Query Match 63.2%; Score 12; DB 3; Length 9190;
Best Local Similarity 100.0%; Pred. No. 1.18e+00;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 5613 ttacaatat 5624
|||||
|TTACAAATATTA 13
Oy 2 TTACAAATATTA 13

```

```

RESULT 6 HIVRCSF 9540 bp ss-RNA VRL 01-DEC-1988
LOCUS Human Immunodeficiency virus type 1, isolate JRCSF; complete
DEFINITION genome.
ACCESSION M38429
SOURCE HIV-1 proviral DNA from extracellular virus taken from cerebral
spinal fluid (1986). Infectious clone.
REFERENCE 1 (bases 1 to 9540).
AUTHORS Koyanagi, S., Chen, I.S.Y. et al.
JOURNAL Unpublished (1988).
STANDARD
COMMENT

```

Kindly provided in computer-readable form by Irvin Chen, UCLA School of Medicine, Los Angeles. JRCSF and JRFL (see <HIVJRL> were isolated from cerebral spinal fluid and brain tissue of the patient JR, who died with Kaposi's sarcoma and severe AIDS encephalopathy (Science 236, 819-822, 1987). Both clones are infectious, but JRFL productively infects macrophages while JRCSF does not. (Peripheral blood was not available from the patient).

```

FEATURES
pept < 2085 5108
pept 790 2304
pept 5108
description gag polyprotein (NH2-terminus uncertain; AA at 2085)
vif protein
vpr protein
tat protein, exon 2 (first expressed exon)
tat protein, exon 3 (AA at 8367)
rev protein, exon 2 (first expressed exon)
rev protein, exon 3 (AA at 8368)
vpu protein
envlope polyprotein
nef protein
3' LTR
5' LTR
R repeat 5' copy
SP1 binding site III
SP1 binding site II
SP1 binding site I
primer (Lys-tRNA) binding site
3425 a 1691 c 2308 g 2116 t
ORIGIN 5' terminus of 5'LTR.

```

```

Query Match 63.2%; Score 12; DB 3; Length 9540;
Best Local Similarity 100.0%; Pred. No. 1.18e+00;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 6082 ttacaatat 6093
|||||
|TTACAAATATTA 13
Oy 2 TTACAAATATTA 13

```

```

RESULT 7 HIVSF2 9737 bp ss-RNA VRL 01-MAR-1989
LOCUS

```



DEFINITION Human immunodeficiency virus type 1, isolate ARV-2/SF2, complete proviral genome.

ACCESSION K02007

SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate ARV-2, integrated proviral DNA and circular unintegrated DNA from human cells; clones lambda-[7A,7D,8A,8B,9B]. Infectious clone.

REFERENCE 1 (bases 1 to 9737) Sanchez-Pescador, R., Power, M.D., Barr, P.J., Steimer, K.S., Stempien, M.M., Brown-Shimer, S.L., Gee, W.W., Renard, A., Randolph, A., Levy, J.A., Dina, D. and Luciw, P.A. Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2) Science 227, 484-492 (1985)

JOURNAL STANDARD full staff\_review

COMMENT 5 variations were observed between the clones used for sequencing:

1) "g" occurs at position 332 in 7A and 9B; "a" occurs in 8A.  
 2) "g" occurs at position 340 in 7A and 9B; "a" occurs in 8A.  
 3) "a" occurs at position 4233 in 8A; "g" occurs in 8B.  
 4) "t" occurs at position 4677 in 8A; "c" occurs in 8B.  
 5) "a" occurs at position 6215 in 7D; "g" occurs in 7A and 9B

FEATURES

CDS <2092..5103 location/Qualifiers

join(5840..6054,8384..8474)  
 /note="pol polyprotein (NH2-terminus uncertain; AA at 2092)"  
 /codon\_start=2092

CDS join(5840..6054,8384..8474)  
 /note="tat protein"  
 /codon\_start=5840  
 join(5979..6054,8384..8658)  
 /note="rev protein"  
 /codon\_start=5979

LTR 2..635  
 /note="5' LTR"

variation 332..332  
 /note="g in 7A, 9B; a in 8A"

variation 340..340  
 /note="g in 7A, 9B; a in 8A"

misc\_binding 378..387  
 /note="Spl binding site III"

misc\_binding 389..398  
 /note="Spl binding site II"

misc\_binding 400..409  
 /note="Spl binding site I"

repeat\_region 454..551  
 /note="R repeat 5' copy"

prim\_transcript 455..9652  
 /note="genomic mRNA"

prim\_transcript 455..9652  
 /note="tat, rev, nef subgenomic mRNA"

misc\_binding 637..654  
 /note="primer (Lys-tRNA) binding site"

intron 745..5783  
 /note="tat, rev, nef subgenomic mRNA intron 1"

CDS 791..2299  
 /note="gag polyprotein"

variation 4233..4233  
 /codon\_start=791

variation 4677..4677  
 /note="a in 8A; g in 8B"

variation 5048..5626  
 /note="t in 8A; c in 8B"

CDS 5048..5626  
 /note="vif protein"

CDS 5366..5859  
 /codon\_start=5048

CDS 5840..6054  
 /note="vpr protein"

exon 5840..6054  
 /codon\_start=5566

exon 5979..6054  
 /note="tat protein, exon 2 (first expressed exon)"

intron 6055..8383  
 /note="rev protein, exon 2 (first expressed exon)"

intron /note="tat cds intron 2"  
 6055..8383

intron /note="rev cds intron 2"  
 6055..8383

CDS /note="tat, rev, nef subgenomic mRNA intron 2"  
 6071..6187

variation /note="vpu protein"  
 /codon\_start=6071  
 6215..6215

CDS /note="a in 7D; g in 7A, 9B"  
 6233..8800

exon /note="envelope polypolyprotein"  
 /codon\_start=6233

exon 8384..8474

exon /note="tat protein, exon 3 (AA at 8385)"  
 8384..8658

CDS /note="rev protein, exon 3 (AA at 8386)"  
 8802..9434

LTR /note="nef protein"  
 /codon\_start=8802  
 9103..9736

repeat\_region /note="3' LTR"  
 9555..9652

misc\_signal /note="R repeat 3' copy"  
 9628..9633

BASE COUNT 3445 a 1738 c 2377 g 2177 t

ORIGIN 5' terminus of LTR-03.

Query Match 63.2%; Score 12; DB 3; Length 9737;  
 Best Local Similarity 100.0%; Pred. No. 1.18e+00;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6080 ttcaaatatatta 6091

QY 2 ttcaaatatatta 13

RESULT 8

LOCUS SIVAGMT49 1773 bp ds-DNA VRL 09-JUN-1993

DEFINITION Simian immunodeficiency virus, isolate TAN-49, complete gag cds.

ACCESSION L19253

SOURCE Simian immunodeficiency virus proviral DNA, isolate TAN-49 from a Ugandan tantalus monkey (Cecopithecus tantalus).

ORGANISM Simian immunodeficiency virus

REFERENCE 1 (bases 1 to 1773) Viridae; ss-RNA enveloped viruses; Positive strand RNA viruses; Retroviridae; Lentivirinae.

AUTHORS Hirsch, V.M., McGann, C., Dapolito, G., Goldstein, S., Ogen-Odoi, A., Biriyawho, B., Lakwo, T. and Johnson, P.R.

JOURNAL Identification of a new subgroup of SIVagm in tantalus monkeys

STANDARD full staff\_review

COMMENT kindly submitted prior to publication by Dr. Vanessa Hirsch, Immunodeficiency Viruses Section, NIAID/NIH, Washington, D.C., Rockville MD 20852. SIVagm-tan strains cluster with other previously characterized agm strains taken from vervets and givets. See related tantalus isolate sequences 17, 27, 40, and 9 (accession numbers L19250-L19252, L19254).

FEATURES

exon Location/Qualifiers

195..1773

gene="gag"

1..1773

source /organism="Simian immunodeficiency virus"

/cell\_line="CEMS"

/haplotype="na"

/proviral

/sequenced\_mol="DNA"

BASE COUNT 594 a 375 c 480 g 324 t

ORIGIN 193 nt upstream from the gag cds start.

Query Match 57.9%; Score 11; DB 3; Length 1773;  
 Best Local Similarity 100.0%; Pred. No. 7.89e+00;



Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 817 tattaagaag 827  
 |||||||||  
 QY 9 TATTAAGAAG 19

RESULT 9  
 LOCUS HIV2CAM4E 2583 bp ss-RNA VRL 02-AUG-1994  
 DEFINITION Human immunodeficiency virus type 2 gp160 envelope sequence,  
 isolate HIV2CAM4E, complete cds.  
 ACCESSION U05356  
 SOURCE Human immunodeficiency virus type 2 DNA derived from a patient from  
 Guinea Bissau.  
 ORGANISM Human immunodeficiency virus type 2  
 Viridae: ssRNA enveloped viruses; Positive strand RNA virus;  
 Retroviridae: lentivirinae;  
 1 (bases 1 to 2583)  
 Breuer,J., Douglas,N.W. and Daniels,R.S.  
 Sequence analysis of env genes of ten HIV-2 isolates and potential  
 identification of domains susceptible to cross neutralizing antibodies  
 JOURNAL Unpublished  
 STANDARD full staff review  
 COMMENT This envelope sequence is one of a group of six isolated from  
 patients in Guinea-Bissau:  
 HIV2CAM1E (U05359) HIV2CAM2E (U05354) HIV2CAM3E (U05355)  
 HIV2CAM4E (U05356) HIV2CAM5E (U05357) HIV2CAM6E (U05358)  
 These sequences cluster with HIV-2 subtype A sequences.  
 FEATURES  
 source  
 1..2583  
 /isolate="HIV2CAM4"  
 /lab\_host="Human T-cell line"  
 /organism="Human immunodeficiency virus type 2"  
 /proviral  
 1..2583  
 /gene="envelope (env)"  
 /note="NCBI gi: 520432"  
 /codon\_start=1  
 /product="gp160"  
 /translation="MAGCKSQLIALLITSLIYCAQYVYVFGIPAMRNASIPFLC  
 /ATNRDTWGTICLPDNDYOEIPLNTEAFAPANNITTEOIEVNNLFEISIKPV  
 KILPLCVAMKCKRMGTNTTATTAATSTLITNDLTNLTSTGANGNCGLEFEELMPC  
 QPNMTGLORDKRYRNETVYSQDYCDTRNTNTKTDNSTCIYMHCHTSTYTESCDK  
 HIWDTMRFRYCAPGPFALRCNDYDSEKPCVCSVVAASCTRMETSTWFGENG  
 RAENRTYIYMHCKNDRITISLTHNLSVHCSPKNTVPTTLMSGLVFSQPIHMR  
 PROAMCFKGEWKGAKKEVQTLLEHPRKGTATNTNTRIAEPGSDPEVAYMTNC  
 RGEFLYCNMTWFLNENRTGPGHNYVCHROIILNTHMKVGNVYLPREGELVCN  
 STVSLINIMRNDSTNTPESAELVRLGIDYKLIETVPIGFAFPTDOKRSBPAP  
 TRSKGVYVLFGLFLAAGTAMGTASLITLSKQSTRLAGIYQOQQLDYVERQEM  
 LRLTWGTRKNLQARVTALEKYLKDDAQLNSWGCVRQVCHTTPVANDSLTPVNNMT  
 QWELRLTRYLDANISQLEQAQIQEKNMYELOLNSMDVFSNFDLSWRYIQYG  
 VYIVGIVLRVIGVYIOMISRLKRGYRPFSSPGYLOQIHIDKQAPAGEEED  
 VGDVNGDSLMPWPLAYLHFLRLRLRLRLRLTSLSCRDLSRFPLQALVALRDMLR  
 LKAAIYQGCWIEIOAFQAFATARETLAGAMRNLMEMGRGRIILAVPRIRQGAEL  
 LALL"

BASE COUNT 837 a 533 c 616 g 597 t  
 ORIGIN envelope cds start

Query Match 57.9%; Score 11; DB 3; Length 2583;  
 Best Local Similarity 100.0%; Pred. No. 7.89e+00;  
 Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 1398 tacaatatla 1408  
 |||||||||  
 QY 3 TACAATATTA 13

RESULT 10  
 LOCUS HIV14575 3120 bp ds-DNA VRL 08-JUN-1994  
 DEFINITION Human immunodeficiency virus type 1, US3 envelope glycoprotein (env)  
 gene, tat gene, rev gene, vpu gene, and nef gene.  
 L14575  
 ACCESSION L14575  
 KEYWORDS env gene; envelope glycoprotein; nef gene; rev gene; tat gene;

SOURCE  
 ORGANISM Human immunodeficiency virus type 1 blood DNA.  
 Human immunodeficiency virus type 1  
 Viridae: ss-RNA enveloped viruses; Positive strand RNA virus;  
 Retroviridae: lentivirinae.  
 1 (bases 1 to 3120)  
 Mascola,J.J., Louwagie,J.J., McCutchan,F.E., Fischer,C.L.,  
 Hegerich,P.A., Wagner,K.F., Fowler,A.K., McNeil,J.G. and Burke,D.S.  
 Two antigenically distinct subtypes of human immunodeficiency virus  
 type 1: viral genotype predicts neutralization serotype  
 J. Infect. Dis. 169, 48-54 (1993)  
 full automatic  
 NCBI gi: 401758  
 FEATURES  
 source  
 1..3120  
 /organism="Human immunodeficiency virus type 1"  
 /cell\_type="lymphocyte"  
 /proviral  
 /sequenced\_mol="DNA"  
 /tissue\_type="blood"

BASE COUNT 1095 a 529 c 754 g 742 t  
 ORIGIN  
 Query Match 57.9%; Score 11; DB 3; Length 3120;  
 Best Local Similarity 100.0%; Pred. No. 7.89e+00;  
 Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 97 tacaatatla 107  
 |||||||||  
 QY 3 TACAATATTA 13

RESULT 11  
 LOCUS SIVAGMTYO 9170 bp ss-RNA VRL 01-MAR-1989  
 DEFINITION Simian (African green monkey) immunodeficiency virus, isolate  
 TYO-1, complete genome.  
 X07805  
 Simian (African green monkey) immunodeficiency virus, extra-  
 chromosomal closed-circular DNA, TYO-1 clone lambda-SAH12.  
 1 (bases 1 to 9170)  
 Fukasawa,M., Mura,T., Hasegawa,A., Morikawa,S., Tsujimoto,H.,  
 Miki,K., Kitamura,T. and Hayami,M.  
 Sequence of simian immunodeficiency virus from African green  
 monkey, a new member of HIV/SIV group  
 Nature 333, 457-461 (1988)  
 JOURNAL  
 STANDARD  
 COMMENT full staff review

This sequence was obtained from a lambda clone derived from an  
 African green monkey imported from Kenya.

SIVAGMTYO appears to lack a YPR coding region. [1] suggests that  
 this may account for SIVAGMTYO's lack of pathogenicity. The env  
 cds contains an in-frame stop codon at positions 8077-8079, similar  
 to the in-frame stop codons found in the SIV macaque and HIV-2  
 ROD35 envelope coding regions, except that SIVAGMTYO's stop codon  
 is found 96bp downstream from the 3' splice junction of tat and  
 rev, which is the location of the stop codon in the other  
 sequences.

[1] finds that SIVAGMTYO is approximately equally distantly related  
 to the HIV-1's and the HIV-2/SIV (macaque) group; over a  
 "conserved" stretch of the pol protein, TYO is only 87% similar to  
 AGM385 and 82% similar to AGM265, the latter two differing by 20%.  
 Location/Qualifiers  
 <1..219  
 /note="5' LTR"  
 <1634..4819  
 /note="pol polypeptide (NH2-terminus uncertain; AA at  
 1634)"  
 /codon\_start=1634  
 join(5546..5763,7984..8068)  
 /note="tat protein"  
 /codon\_start=5546  
 join(5703..5763,7984..8177)

FEATURES  
 LTR  
 CDS  
 CDS  
 CDS



/note="rev protein"  
/codon\_start=5703  
prim\_transcript 1..9170  
/note="genomic mRNA"  
prim\_transcript 1..9170  
/note="tat, rev, nef subgenomic mRNA"  
repeat\_region 1..117  
/note="R repeat 5' copy"  
220..237  
misc\_binding /note="primer (Lys-tRNA) binding site"  
432..1991  
CDS /note="gag polyprotein"  
/codon\_start=432  
4755..5462  
/note="vif protein"  
/codon\_start=4755  
5236..5595  
/note="vpx protein"  
/codon\_start=5236  
5546..5763  
exon /note="tat protein, exon 2 (first expressed exon)"  
5703..5763  
/note="rev protein, exon 2 (first expressed exon)"  
5764..7983  
intron /note="tat intron 2"  
5764..7983  
/note="rev intron 2"  
5764..7983  
intron /note="tat, rev, nef subgenomic mRNA intron 2"  
5770..8367  
CDS /pseudo  
/note="env polyprotein (in-frame stop codon at 8077)"  
/codon\_start=5770  
7984..8068  
exon /note="tat protein, exon 3 (AA at 7985)"  
7984..8177  
exon /note="rev protein, exon 3 (AA at 7986)"  
8077..8079  
misc\_feature /note="in-frame stop in env cds"  
8201..8890  
CDS /note="nef protein"  
/codon\_start=8201  
8547..>9170  
LTR /note="3' LTR"  
9054..9170  
repeat\_region /note="R repeat 3' copy"  
9151..9156  
misc\_signal /note="mRNA polyadenylation signal"  
BASE COUNT 3172 a 1720 c 2263 g 2015 t  
ORIGIN Cap site of genomic RNA.  
Query Match 57.9% Score 11; DB 3; Length 9170;  
Best Local Similarity 100.0%; Pred. No. 7.89e+00;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 5449 tacaatattat 5459  
|||||  
QY 3 TACAATATTAT 13  
RESULT 12  
LOCUS SIVSABIC 10036 bp ss-RNA VRL 25-MAY-1994  
DEFINITION Simian immunodeficiency virus (SIVagm) complete genomic sequence,  
African green monkey isolate SIVagmSAB-1.  
ACCESSION U04005  
SOURCE Simian immunodeficiency virus PCR-amplified DNA extracted from  
cocultured PBMCs derived from a wild female asymptomatic sabaues  
African green monkey captured in its natural habitat of Senegal.  
ORGANISM Simian immunodeficiency virus  
Virusidae; ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
REFERENCE 1 (bases 1 to 10036)

AUTHORS Jin, M.-J., Hui, H., Robertson, D. L., Mueller, M. C., Barre-Sinoussi, F.,  
Hirsch, V. M., Allan, J. S., Shaw, G. M., Sharp, P. M., and Hahn, B. H.  
TITLE Mosaic Genome Structure of Simian Immunodeficiency Virus from West  
African Green Monkeys  
JOURNAL EMBO J 13, 2935-2947 (1994)  
STANDARD full staff review  
COMMENT This sequence was obtained as part of a study designed to determine  
the breadth of genetic variation of SIV infecting African green  
monkeys. With this aim in mind, Jin et al. PCR-amplified, cloned  
and sequenced the LTR and partial envelope region from 12 isolates  
(ver-1, ver-2, gri-2, gri-3, sab-1, sab-2, sab-3, sab-4, tan-1,  
tan-17, tan-40, and tan-49). These isolates were derived from  
different African green monkey species inhabiting various  
geographical regions.  
Virus sab-1 was derived from a wild female asymptomatic sabaues  
African green monkey. She was caught in her natural habitat of  
Senegal and subsequently exported to the United States. From the  
point of capture and throughout the study, she was housed in an  
individual cage.  
Phylogenetic analysis suggests that each AGM species is infected  
with a distinct subtype of SIVagm. Interestingly, the sabaues viral  
genome seems to be a mosaic of many viruses infecting different  
primate species [1]. Virus sab-1 forms a distinct lineage in the 5'  
end of gag, clusters closest to the HIV2/SIVsm group in the 3' gag  
and 5' pol regions, and clusters closest to the other SIVagm  
lineages in the 3' pol, env and nef regions [1]. This feature  
was demonstrated for two viral sabaues strains, sab-1 and sab-037  
(sequenced only over the 5' end of the pol gene). Jin et al.  
conclude that an ancestor of the sabaues virus was a recombinant  
with two breakpoints: one roughly at codon 510 of gag and the other  
at 710 of pol.  
NCBI gi: 466229  
FEATURES  
source location/Qualifiers  
1..10036  
/clone="SAB-1/MJ8"  
/strain="SIVagmSAB-1"  
/organism="Simian immunodeficiency virus"  
/specific\_host="African green monkey"  
/provisional  
/note="sabaues subtype"  
1..767  
LTR 481..527  
/note="putative TAR 1"  
528..568  
LTR /note="putative TAR 2"  
1017..2681  
CDS /note="NCBI gi: 466230"  
/codon\_start=1  
/product="GAG polyprotein"  
/translation="MGASNSVLSGKRLAFESVRRLRPNPKKKYKLRHLVMAKSELDRE  
SLSENVLEKGVYLLSVLLPIVPTGSENLIALFCVACHAEIYKVDTEAKA  
KYKEVYPAMTESATATSSGOTKELOAKKNEPTVTPGGSGSNITVSYNNVHOPL  
SPRTLNAMVKVLEEKESAEVVPMSALAEGLIPDIOMLVAEHOGLQIVDVI  
NEADMDLRHPPOOPPOAGVLRPOSDIGTSTIOEQLIEMTTRRANVNGNII  
KGMILGLQKCYKMYNPVNIIDIKOGPEPEFDYRFEKALRQOTPOAVNWTOS  
LIIONANPDKTVLKGIGMNPTELEMLTACOGSIGAOKHAKARARROGCKGCSPPDMK  
IFVOCARPGRPIGRGRBPINPTICYNCKRGKHLARCKARROGCKGCSPPDMK  
DCQOVNFTLGRGWRGRKPRNPELTSIRTPAPWRKDSRPENNTAURPPTRGPPD  
DPAIALLQYAVGKRQKQWONHSPQSPYEAVYSLSLFGEDQ"  
42387..5506  
/note="N-terminal uncertain; NCBI gi: 466231"  
/codon\_start=1  
/product="POL protein"  
/translation="FFRYVPPLUGQRTQEPFSDLRHTNSPNRTGLQAGKIVKQTS  
DQRTARRSSNSPVAVCCSGEATFAAKPLATEPRLGGLQFQVSLWRPMKTVY  
LEGQVYALDLGADSVYIOGIELDDNNKPRIIGIGICINVKVANOVEIEDTKK  
AIIIGEPVNIIGRNVLQGLVTLNOREIEPIKVLKPOGDPRIROMPLSKKI  
BALKICEDLEKGLERIGENPNYNTVPIIRKDKTQWRIMREOLNKGSTODFOE  
VQIGIPHAGLQOQROIVLIDIGAFESCPLDPDOKYTAFTIPSVNNREGIOFYK  
VLPQGMKSPRTFOTANKIIOEFKQKPPVDIYIMDMILASRPRAEHLVYQOL  
RDYLETWQKTIPEKFFQDPPLVMGYELYPKRWQLOITLPEREVTVNDIQKLVGK  
LMAQSIYTGIRTKHLRLIRGARPLTIYQWTEBAEDELLENROILROKQOQGYDP  
ALPLRAKVLKLDGQMGQYIYQPEKKILKVGKAYAKIKTAHTNELMAGLVQKIGES



IVYQIPIIMELPERELMEQMSDYQVYTIPEWENSTQQLRLMKYLVKDPICE  
AVIYVDGANNRNSKRGAGYLDTRGDQVVALENTNKALEALLALRLSGSKVNI  
ITDSQYAMGIIAGEPTESDNNIYVOIIEELIKKEAVYIANYAHGNGNEIDKIVS  
OGIRQVFLDLRIEAEQEHDRKYHANRMRQEGFIPAVAKEIYACKCOIKESVH  
GQVAPSFGWQMDCTHEGKIIIVYHVASSEIEVEYPOEGKATAEKILISRP  
ITOLHTONGTSTQOVVALICOMGKEHTPEPVYNOGYVESNMOKLEKIIQIR  
DABREIYAVIAVHINHRKRGIGISAEARLNIITELETRLOOKISKIONERY  
YVREGDPWAKPLIKWGEAVVIOEGELKTIIPRRKATINDYGRALDSQAPLEG  
NRTAGEVD"  
5439..6143  
/note="NCBI gi: 466232"  
/codon\_start=1  
/product="vif protein"  
/translation="MEKHIIVRPLMKVTGGQERMTSLVKYHMHVSKOCVHRYTPHT  
KIRNMWYSYQEMVPIPLKDALIKVYHMLPEKMGLETYATIGISGSEWETLDPW  
TADHIIHWSYFPCEDRAVQOAIKGEKIKHOGKHOGVQISQIYALRYVNGI  
RRVAPTSRSGSSQSGPSQRDRFMANNMGFAQAVARMAPRHVYTGQFGFVPLPK  
ESPPSYVYCGRTSH"  
5911..6333  
/note="NCBI gi: 466235"  
/codon\_start=1  
/product="VPX protein"  
/translation="MASGMLPYGDDPKDPKPNRREIPEIETMDLPREPDEML  
RMDQDENSEAQCRRPLIFRLMNIYERPAIDHGOQLRGLGWTYKICILOKALFVHM  
KGRCKPETHPAIGFAGGPPGLGAGSAGASAPGL"  
6254..6475  
/note="NCBI gi: 466233"  
/codon\_start=1  
/product="TAT protein"  
/translation="MDROEARPOVMELOELHRLDQACNTOFCYKCHLCEH  
KKAIGIRIYVPRPRAKSKISHNOVSLHN"  
Join(6411..6474,8683..8918)  
/note="NCBI gi: 466234"  
/codon\_start=1  
/product="REV protein"  
/translation="MSIGOEELRFRIRIKFLYTNTPPGCGTARORARORAMAO  
ROQYIHLAERILETPVSOIDHLOEFQDLVDNLQPPSLPFGHPTENQIANS"  
6481..8679  
/note="NCBI gi: 466236"  
/codon\_start=1  
/product="ENV protein"  
/translation="MKLLTYLMSGSLVWLYQVYVFGIYPMKNSVQACRTP  
NTNLMASINCLIPDEBECTIAEVLPIPTTEKFDAMKNRNPVAGAESIHILFESTLK  
PCFALSPWICIMANCYRLEGAATTPSTSTARPEVSVGENDVIEDMKKEADAMC  
SFAGAGRDYKKNSTVMDQEVYCEGREGKSNATHGCMYHCHNTSYKEACDRT  
YMDTEFLRYCAPAGYALRCADDTYSGHAKRNTVSACTRLINTVSTGIGINSYV  
ANREIEMOKNNSNDVYIIRLNRYFNILRCRPRGNKTVLPITMAGIYVHSOKYNR  
LQAMCMTGGMWRAKMKVETIYRLPKKYSINDIKIFLQROMDPESEFFFPNC  
QGEFFTCAMDVFNILNKNVSDPHNNCAKNNTPCWORTYVPCHIROYVNDWYTLK  
KTYAPREGHLECNSTATALYELNENSKNTNVTLSPOESIVANEIGDKVLEIKP  
IGFAPTRYRRTEPERORRVFVGLFGLFCAAMAAATALTAVSQOQLAIILO  
QNLIAAEVQOQOQMLKILIMGVKNLNAVITLLEYLEQOALNMGCAFOYCHTTLV  
MKYNTPDWEMTQEMEROIEKYEANSTILBOAHDEOKNLSYKISWSPFMSH  
FDLRKMGKMKIATVYAGITVAVLVYITILKFRKGYAPLSSPSH"  
8921..9601  
/note="NCBI gi: 466237"  
/codon\_start=1  
/product="NEF protein"  
/translation="MGKSSKQOORHSLMISKLROAPVIOYDMLDPLIGSSSHIOE  
ECASLRDGLRGDSSRTBEGYKMKHOGRPSTYDEDEEYGVPRCPLELRMYTK  
LAIDFGHLKKGGLGEGITYSERKKIIDLALWEGIVDMONYTDPGRTYKPCFG  
WCFRLVPVDSLSEAKNSNCLLHPAQAAYDDAMKETLVKFPDLLAVDVAVMLHP  
EOWPSAOG"  
9270..10036  
LTR  
BASE COUNT 3293 a 1975 c 2517 g 2251 t  
ORIGIN

Query Match 57.9%; Score 11; DB 3; Length 10036;  
Best Local Similarity 100.0%; Pred. No. 7.89e+00;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2257 caaatattaaa 2267  
|||||

OY 5 CAAATATTAAA 15

RESULT 13  
LOCUS STYTAN40E 1119 bp ss-RNA VRL 25-MAY-1994  
DEFINITION Simian immunodeficiency virus (Styagm) partial envelope sequence,  
African green monkey isolate StYagm4N-40.  
ACCESSION U04001  
SOURCE Simian immunodeficiency virus PCR-amplified DNA extracted from  
cocultured PMCs derived from a wild asymptomatic tantalus African  
green monkey captured in its natural habitat of Uganda.  
ORGANISM Simian immunodeficiency virus  
Virus; ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
REFERENCE 1 (bases 1 to 1119)  
Jin,M.J., Hui,H., Robertson,D.L., Mueller,M.C., Barre-Sinoussi,F.,  
Hirsch,V.M., Allan,J.S., Shaw,G.M., Sharp,P.M. and Hahn,B.H.  
Mosaic Genome Structure of Simian Immunodeficiency Virus from West  
African Green Monkeys  
EMBO J. 13, 2935-2947 (1994)  
COMMENT full staff review  
This sequence was obtained as part of a study designed to determine  
the breadth of genetic variation of SIV infecting African green  
monkeys. With this aim in mind, Jin et al. PCR-amplified, cloned  
and sequenced the LTR and partial envelope region from 12 isolates  
(ver-1, ver-2, gri-2, gri-3, sab-1, sab-2, sab-3, sab-4, tan-1,  
tan-17, tan40, and tan-49). These isolates were derived from  
different African green monkey species inhabiting various  
geographical regions.  
Virus tan-40 was derived from a wild asymptomatic tantalus  
African green monkey. This monkey was caught in its natural  
habitat of Uganda.  
NCBI gi: 466221

FEATURES  
source : location/Qualifiers  
1..1119  
/clone="envTAN40"  
/strain="StYagm4N40"  
/organism="simian immunodeficiency virus"  
/specific\_host="African green monkey"  
/note="tantalus subtype"  
<1..1119  
/note="NCBI gi: 466222"  
/codon\_start=1  
/product="envelope protein"  
/translation="TLRLRCAPAGFLLKCRDYNVAGFCINNVSVSTGLMNTTI  
SSAFGLGSKAENRTEIMWKGNSNDVYIIRKHKHOLKICIRPGKKTLYLPVIMAG  
LVHESQVYKNAOAMCHFCGDMKGNKKEVRIYTLPRKYGCTKNTPEIMLRQMG  
DPEANINWNCQGEFFTCYTDPMWVNNLNNESDSGKSYNVEGNTCTPEKDKCRRTY  
ACHIRSVNWDVILSKRTYAPREGHLECNSTATALYELNENSKNTNVTLSPOESIVANEIGDKVLEIKP  
IGFAPTRYRRTEPERORRVFVGLFGLFCAAMAAATALTAVSQOQLAIILO  
QNLIAAEVQOQOQMLKILIMGVKNLNAVITLLEYLEQOALNMGCAFOYCHTTLV  
MKYNTPDWEMTQEMEROIEKYEANSTILBOAHDEOKNLSYKISWSPFMSH  
FDLRKMGKMKIATVYAGITVAVLVYITILKFRKGYAPLSSPSH"  
ALTYVSOQLAGLIQOQOQMLKILIMGVKNLNAVITLLEYLEQOALNMGCAFOYCHTTLV  
MKYNTPDWEMTQEMEROIEKYEANSTILBOAHDEOKNLSYKISWSPFMSH"  
BASE COUNT 383 a 194 c 268 g 274 t  
ORIGIN approximately 727 bp downstream from beginning of env cds

Query Match 52.6%; Score 10; DB 3; Length 1119;  
Best Local Similarity 100.0%; Pred. No. 4.19e+01;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 78 tattgttaac 87  
Cp 10 TATTGTAAAC I

RESULT 14  
LOCUS HUMMS1 1478 bp ss-mRNA PRI 10-JUN-1992  
DEFINITION Human mRNA for the M5S1 cellular factor, complete cds.  
ACCESSION D11094  
SOURCE Human cDNA clone M5S1 (from HeLa cDNA library)  
ORGANISM Homo sapiens  
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;  
Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.  
REFERENCE 1 (bases 1 to 1478)  
Shibuya,H., Irie,K., Ninomiya-Tsuji,J., Goebel,M.G., Taniguchi,T.



TITLE and Matsumoto, K.  
A novel human gene encoding a positive modulator of HIV  
Tat-mediated transactivation  
JOURNAL Nature 357, 700-702 (1992)  
STANDARD full staff review  
COMMENT These data kindly submitted in computer readable form by Kenji Irie, Department of Molecular Biology, Faculty of Science, Nagoya University, Chikusa-ku, Nagoya 464-01, Japan. The MSI protein is reported to be 42% similar to the human TRP-1 protein, which binds the HIV tat protein in vitro. Endogenous MSI enhances tat-mediated transactivation, leading the authors of [1] to conclude that MSI has a "key role in activation of HIV genes regulated by tat."

FEATURES  
CDS  
Location/Qualifiers  
67..1368  
/function="positive modulator of HIV Tat-mediated  
transactivation"  
/product="MSI protein"  
/gene="MSI"  
/codon.start=1  
/translation="MPYLGADQKTEDEKDKPIRALDEGDALTKYGSYRSRQ  
IKQVEDIIOQLIKKINELTGKESDGLAPALADAKOTLOSQPLQVABRTKI  
NABSEDPKTIINQAKFVYDSDQYAPTDIEGRVGVDRNKYQHLPPIRDT  
VTMMQVEKPDVYSDVGGEQIEKLEVEYETPLHPRVNGIIEPPGVLLFEP  
GTGKTLGARAVANTDACEFRVIGSELVQKYGAGAMVRELFEAMATKACLIFFE  
IDAIGARFDGAGDNQVORTMELINQDGFDPKCNKIVMATNRPDTLDPALMP  
GLDRKIEFSLPDEGRTHIFKIHARMSVSRDIRFELIACNSGAEISVCTEA  
GMAIRARRKATKDFLEAVNKIKYKAFSAIPRMTYN"

source  
1..1478  
/organism="Homo sapiens"  
/sequenced\_mol="cDNA to mRNA"  
BASE COUNT 439 a 287 c 375 g 377 t  
ORIGIN 67 nt upstream from cds start.

Query Match 52.6%; Score 10; DB 1; Length 1478;  
Best Local Similarity 100.0%; Pred. No. 4.19e+01;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 258 tattaagaa 267  
|||||  
OY 9 TATTAAAGA 18

RESULT 15  
LOCUS HIV2D808 1512 bp ss-RNA VRL 15-Apr-1994  
DEFINITION Human immunodeficiency virus type 2 (HIV-2) gp105 envelope  
glycoprotein sequence, sample D808, partial cds.  
ACCESSION U07107  
SOURCE Human immunodeficiency virus type 2 from PCR-amplified DNA extracted  
from uncultured PBMCs derived from a patient from Bombay, India whose  
sera was dually reactive (HIV/HIV2).  
ORGANISM Human immunodeficiency virus type 2  
Virus: ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
1 (bases 1 to 1512)  
Grez, M., Dietrich, U., Balfe, P., Von Briesen, H., Maniar, J. K.,  
Mahamre, G., Delwart, E. L., Mullins, J. I. and Ruebsamen-Watigmann, H.  
Genetic analysis of human immunodeficiency virus type 1 and 2  
(HIV-1 and HIV-2) mixed infections in India reveals a recent spread  
of HIV-1 and HIV-2 from a single ancestor for each of these viruses  
J. Virol. 68, 2161-2168 (1994)  
JOURNAL full staff review  
STANDARD This sequence was obtained as part of a study to determine the  
genetic breadth of HIV-2 variation in India. In this study, blood  
samples were obtained from four patients from Bombay, India (D766,  
D808, D868, and D1024) and from one patient from Goa, India (D1071);  
each of the five patients was serologically dually reactive. The  
mean genetic divergence of the set of four sequences from Bombay was  
7.3%, while the sequence from Goa was from 7 to 10.5% different from  
each of the Bombay sequences. The entire set of Indian sequences,  
including both the Bombay and Goa sequences, diverged from eleven  
non-Indian HIV-2 sequences by 23%. Because of the similarity of the  
Indian sequences, Grez et al. suggest that similar HIV-2 strains

might be found elsewhere in India. All five of these sequences  
clustered with HIV-2 subtype A viruses. HIV-1 env sequences from  
Indian patients D766, D808, D868, and D1024 clustered with HIV-1 C  
viruses, and as a set showed a similar degree of interpatient  
genetic similarity as did their HIV-2 counterparts.

FEATURES  
source  
Location/Qualifiers  
1..1512  
/strain="hiv2d808"  
/organism="Human immunodeficiency virus type 2"  
/proviral  
1..>1512  
/gene="env"  
/note="NCBI gi: 471754"  
/codon.start=1  
/function="envelope protein"

CDS  
/product="gp105"  
/translation="MAPQTQLILSILTSACILYCEQVTVYVYALPAMFNASIPLEFC  
ATNRDQWGTIOCLPNDYQOETALNVEAFDAMNNTVTEQALQEDVWNLFEETSKPCV  
KLPIPLCYAMNSKNYNTSAGNTTPTAESMINENSSCIANNCSGLEBEIVNCOFN  
MAGLPDRKPKVNEETWSRDVCCENNPTONRCYMNHCNTSVTSECEKRYMDIRP  
YCTPPGAILKCDNTNSGPEPCSKVVAATRTMMETORTSTMEGFGTAENRTYI  
WSEKNDRTIISLTHYNLSTYCKRPGKTYVPTLMSGIVPHSQPINRRQAMCFK  
GRKKEAQVEKETLTHPRYRGINDTKINFAAPGKSDPEAAVMWNTGCEFLYCM  
TWELMWENTTAQTMHNVPCHIKOIIINWAVGKRVLPREGELTNSVTSLIA  
NIDGQNTITFSAEYQELRLIGIGYKIVETISIGFAPTSERKSSAHRHRE"

BASE COUNT 536 a 304 c 319 g 353 t  
ORIGIN envelope cds start  
Query Match 52.6%; Score 10; DB 3; Length 1512;  
Best Local Similarity 100.0%; Pred. No. 4.19e+01;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1372 acaaatatta 1381  
|||||  
OY 4 ACAATATTA 13

Search completed: Sat Nov 27 15:25:36 1999  
Job time : 9 secs.



**THIS PAGE BLANK (USPTO)**





Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPerch\_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 15:24:13 1999; MasPar time 21.10 Seconds

Tabular output not generated. 192.971 Million cell updates/sec

Title: >US-09-103-287-6  
Description: (1-19) from US09103287.seq  
Perfect Score: 19  
N.A. Sequence: 1 GTTACAAATATTAAAGAG 19  
Comp: CAACTGTTAATATTCCTC

Scoring table: TABLE jmetric  
Gap 60

Mmatch STD : Dbase 0; Query 0

Searched: 271905 segs, 107135622 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

n-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39 40:part40 41:part41 42:part42 43:part43  
44:part44 45:part45 46:part46 47:part47 48:part48  
49:part49 50:part50 51:part51 52:part52 53:part53  
54:part54 55:part55 56:part56 57:part57 58:part58  
59:part59 60:part60

Statistics: Mean 4.586; Variance 2.920; scale 1.570

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description	Pred. No.
1	100.0	19	60	V99652	Murc polynucleotides	9.94e-02
2	100.0	19	48	V53479	DNA encoding a Staphy	9.94e-02
3	100.0	19	60	V80065	Partial nucleotide se	9.94e-02
4	100.0	19	60	V99650	UDP-N-acetylmuramate	9.94e-02
5	100.0	19	58	V74703	Staphylococcus aureus	9.94e-02
6	84.2	16	22	T22923	Human gene signature	3.98e+00
7	84.2	16	8	Q60017	Human brain expressed	3.98e+00
8	84.2	16	31	T75076	DNA encoding aspartyl	3.98e+00
9	84.2	16	58	V74671	Staphylococcus aureus	3.98e+00

Result ID	Score	Query Match	Length DB	ID	Description	Pred. No.
10	78.9	15	30	T40041	Presentillin-1 exon 12	1.31e+01
11	78.9	15	41	T96670	Human presentillin-1 ge	1.31e+01
12	78.9	15	34	T73866	Cotton fibre promoter	1.31e+01
13	78.9	15	58	V74407	Staphylococcus aureus	1.31e+01
14	73.7	14	19	T19814	Human gene signature	4.20e+01
15	73.7	14	53	V89012	EST clone 19942	4.20e+01
16	73.7	14	19	T20074	Human gene signature	4.20e+01
17	73.7	14	48	T98536	DNA encoding a S. pne	4.20e+01
18	73.7	14	1257	N70394	Sequence encoding hum	4.20e+01
19	73.7	14	2027	V07477	Human protease nexin	4.20e+01
20	73.7	14	2028	V07476	Human protease nexin	4.20e+01
21	73.7	14	15	O89924	Protease Nexin-1 type	4.20e+01
22	73.7	14	2029	N81281	Sequence of coding re	4.20e+01
23	73.7	14	2029	Q20494	PN-I alpha analogue,	4.20e+01
24	73.7	14	2029	Q20496	PN-I alpha analogue,	4.20e+01
25	73.7	14	2031	Q89925	Protease Nexin-1 type	4.20e+01
26	73.7	14	2031	N81282	Sequence of coding re	4.20e+01
27	73.7	14	2032	Q22585	PN-I beta analogue, P	4.20e+01
28	73.7	14	2032	Q22586	PN-I beta analogue, P	4.20e+01
29	73.7	14	2032	Q22587	PN-I beta analogue, P	4.20e+01
30	73.7	14	2199	V82778	Clone bp783_3 isolate	4.20e+01
31	73.7	14	58	V74500	Staphylococcus aureus	4.20e+01
32	73.7	14	60	X13040	Staphylococcus faecalis	4.20e+01
33	73.7	14	5635	V52134	Streptococcus pneumoniae	4.20e+01
34	73.7	14	59	V80182	Nucleic acid sequence	4.20e+01
35	73.7	14	20	T12170	pG4-5-CDK-BP cDNA cl	4.20e+01
36	73.7	14	60	X13118	Enterococcus faecalis	4.20e+01
37	73.7	14	57	V83540	PCR-generated region	4.20e+01
38	73.7	14	15109	V74355	Staphylococcus aureus	4.20e+01
39	73.7	14	46	V18273	HRIV-3 FRLH cp45 vac	4.20e+01
40	73.7	14	46	V18272	HRIV-3 US isolate wil	4.20e+01
41	73.7	14	57	V83561	Nucleotide sequence o	4.20e+01
42	73.7	14	57	V83532	Plasmid p3/7(131) enc	4.20e+01
43	73.7	14	57	V83533	Plasmid p3/7(131)26 e	4.20e+01
44	73.7	14	57	V83520	Plasmid p218(131) enc	4.20e+01

#### ALIGNMENTS

RESULT 1  
ID V99652 standard; DNA; 19 BP.  
AC V99652;  
DT 17-MAR-1999 (first entry)  
DE Murc polynucleotides amplifying primer.  
KW Murc gene; UDP-N-acetylmuramate; L-alanine ligase; Murc polypeptide;  
KW bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
KW immunogen; drug; genetic immunisation; PCR primer; ss.  
OS Synthetic.  
ON Staphylococcus aureus.  
PN EP-889123-A2.  
PD 07-JAN-1999.  
PF 26-JUN-1998; 305064.  
PR 03-JUL-1997; US-052720.  
PA (SMK ) SMITHKLINE BEECHAM CORP.  
PI Burnham MKR, Wallis NG;  
DR MPI: 99-062655/06.  
PT New isolated Murc polypeptide from Staphylococcus aureus and related  
PT nucleic acid - useful in diagnosis, treatment and prevention of  
PT bacterial infections  
PS Disclosure; Page 13; 39pp; English.  
CC The invention relates to a UDP-N-acetylmuramate:L-alanine ligase (Murc  
CC polypeptide) encoded by the S. aureus Murc gene. Host cells containing  
CC an expression system comprising the Murc gene can be used for the  
CC recombinant production of the polypeptide. Agonists or the Murc  
CC polypeptide are used to treat conditions requiring increased activity or  
CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or  
CC competitive polypeptide are useful for inhibiting the polypeptide e.g.  
CC bacterial (especially S. aureus) infections. They are also useful against  
CC Helicobacter pylori infections and related cancers, ulcers and gastritis.  
CC The antibacterial agents are useful to treat in-dwelling devices for  
CC infection prevention or generally as wound treatments to prevent adhesion  
CC of bacteria to matrix proteins. The Murc polypeptide is also useful for



CC diagnosing or prognosing a (susceptibility to) disease, for raising  
 CC antibodies; to identify modulators or specific receptors; in rational  
 CC drug design and as an immunogen for vaccines. The Murc gene sequences are  
 CC useful in antisense/ribozyme therapeutics; to detect mutant Murc gene;  
 CC for chromosomal mapping; to determine bacterial serotype; and for genetic  
 CC immunisation. The present sequence represents a primer used for the PCR  
 CC amplification of the Murc polynucleotides.  
 CC Sequence (Murc ORF) of the Murc gene.  
 SQ Sequence 19 BP; 10 A; 1 C; 3 G; 5 T;

Query Match 100.0%; Score 19; DB 60; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 9.94e-02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 gtacacaatatttaagaag 19  
 ||||||||||||||||  
 QY 1 GTTACAAATATTAAAGAG 19

RESULT 2  
 ID V53479 standard; DNA: 619 BP.  
 AC V53479;  
 DT 30-OCT-1998 (first entry)  
 DE DNA encoding a Staphylococcus aureus protein of unknown function.  
 KM Staphylococcus aureus protein; immune response induction; eye infection;  
 KM antibody production; T-cell immune response; gastrointestinal infection;  
 KM respiratory infection; inhibitor; bacterial infection; cardiac infection;  
 KM central nervous system; kidney infection; urinary tract infection;  
 KM antimicrobial compound identification; broad spectrum antibiotic;  
 KW therapy; ss.  
 OS Staphylococcus aureus.  
 PN EP-841394-A2.  
 PD 13-MAY-1998.  
 PF 24-SEP-1997; 307485.  
 PR 24-SEP-1996; US-027032.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PI Black MT, Burnham MKR, Hodgson JE, Knowles DJC,  
 PI Lomanto M, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,  
 PI Ward JM;  
 DR WPI: 98-252940/23.  
 DR P-PSDB: W77686.  
 PT New nucleic acid sequences from Staphylococcus aureus WCHU29 -  
 PT useful in vaccines and for treatment of bacterial infections of e.g.  
 PT respiratory tract and central nervous system  
 PS Claim 1; Page 146; 390pp; English.  
 CC This sequence encodes a Staphylococcus aureus protein of unknown  
 CC function, and represents a DNA sequence of the invention.  
 CC The DNA sequences were isolated from Staphylococcus aureus WCHU29  
 CC (NCIMB 40771). Host cells containing the DNA sequences are used to  
 CC produce polypeptides or fragments. The proteins are used in the treatment  
 CC of disease, for inducing an immune response by administering them, to  
 CC produce antibody and/or T-cell immune response. Antagonists of the  
 CC proteins are used for the inhibition of bacterial polypeptides.  
 CC Conditions which may be treated include bacterial infections, especially  
 CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,  
 CC urinary tract, skin, bones and joints. The proteins can also be used to  
 CC identify antimicrobial compounds which are broad spectrum antibiotics,  
 CC especially useful in the treatment of H. pylori infection.  
 SQ Sequence 619 BP; 208 A; 117 C; 79 G; 215 T;

Query Match 100.0%; Score 19; DB 48; Length 619;  
 Best Local Similarity 100.0%; Pred. No. 9.94e-02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 470 ctctttaattgttaac 488  
 ||||||||||||||||  
 Cp 19 CTTCTTAATTGTGAAC 1

RESULT 3  
 ID V80065 standard; DNA: 660 BP.  
 AC V80065;

DT 17-MAR-1999 (first entry)  
 DE Partial nucleotide sequence of the Murc gene.  
 KM Murc gene; UDP-N-acetylmutamate:L-alanine ligase; Murc polypeptide;  
 KM bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
 KM immunogen; drug; genetic immunisation; ds.  
 OS Staphylococcus aureus.  
 FH Key Location/Qualifiers  
 FT CDS 2..660  
 FT /\*tag= a  
 FT /product= "partial Murc polypeptide"

EP-889123-A2.  
 PD 07-JAN-1999.  
 PF 26-JUN-1998; 305064.  
 PR 03-JUL-1997; US-052720.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PI (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Burnham MKR, Wallis NG;  
 DR WPI: 99-062655/06.  
 DR P-PSDB: W89199.

PT New isolated Murc polypeptide from Staphylococcus aureus and related  
 PT nucleic acid - useful in diagnosis, treatment and prevention of  
 PT bacterial infections  
 PS Claim 2; Pages 4-5; 39pp; English.  
 CC The invention relates to a UDP-N-acetylmutamate:L-alanine ligase  
 CC (Murc polypeptide) encoded by the S. aureus Murc gene. Host cells  
 CC containing an expression system comprising the Murc gene can be used for  
 CC the recombinant production of the polypeptide. Agonists or the Murc  
 CC polypeptide are used to treat conditions requiring increased activity or  
 CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or  
 CC competitive polypeptide are useful for inhibiting the polypeptide e.g.  
 CC bacterial (especially S. aureus) infections. They are also useful against  
 CC Helicobacter pylori infections and related cancers, ulcers and gastritis.  
 CC The antibacterial agents are useful to treat in-dwelling devices for  
 CC infection prevention or generally as wound treatments to prevent adhesion  
 CC of bacteria to matrix proteins. The Murc polypeptide is also useful for  
 CC diagnosing or prognosing a (susceptibility to) disease, for raising  
 CC antibodies; to identify modulators or specific receptors; in rational  
 CC drug design and as an immunogen for vaccines. The Murc gene sequences are  
 CC useful in antisense/ribozyme therapeutics; to detect mutant Murc gene;  
 CC for chromosomal mapping; to determine bacterial serotype; and for genetic  
 CC immunisation. The present sequence represents a partial nucleotide  
 CC sequence (Murc ORF) of the Murc gene.  
 SQ Sequence 660 BP; 233 A; 83 C; 120 G; 224 T;

Query Match 100.0%; Score 19; DB 60; Length 660;  
 Best Local Similarity 100.0%; Pred. No. 9.94e-02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 179 gtacacaatatttaagaag 197  
 ||||||||||||||||  
 QY 1 GTTACAAATATTAAAGAG 19

RESULT 4  
 ID V99650 standard; DNA: 1351 BP.  
 AC V99650;  
 DT 17-MAR-1999 (first entry)  
 DE UDP-N-acetylmutamate:L-alanine ligase (Murc polypeptide) encoding DNA.  
 KM Murc gene; UDP-N-acetylmutamate:L-alanine ligase; Murc polypeptide;  
 KM bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
 KM immunogen; drug; genetic immunisation; ds.  
 OS Staphylococcus aureus.  
 FH Key Location/Qualifiers  
 FT CDS 22..1335  
 FT /\*tag= a  
 FT /gene= "Murc"  
 FT /product= "UDP-N-acetylmutamate:L-alanine ligase  
 (Murc polypeptide)"  
 EP-889123-A2.  
 PD 07-JAN-1999.  
 PF 26-JUN-1998; 305064.  
 PR 03-JUL-1997; US-052720.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.



PA (SMK) SMTKLINE BEECHAM PLC.  
PI Burnham MKR, Wallis NG;  
DR WP1: 99-062655/06.  
P-BSD8: W87771.  
PT New isolated MucC polypeptide from *Staphylococcus aureus* and related  
PT nucleic acid - useful in diagnosis, treatment and prevention of  
PT bacterial infections  
PS Claim 2: Pages 3-4; 39pp: English.  
CC The present sequence represents a MucC gene encoding a *Staphylococcus aureus* UDP-N-acetylglucosamine-6-phosphate 4-epimerase (MucC polypeptide). Host cell containing an expression system comprising the MucC gene can be used for the recombinant production of the polypeptide. Agonists or the MucC polypeptide are used to treat conditions requiring increased activity or expression of the polypeptide. Antagonists, inhibitory nucleic acid or competitive polypeptide are useful for inhibiting the polypeptide e.g. bacterial (especially *S. aureus*) infections. They are also useful against *Helicobacter pylori* infections and related cancers, ulcers and gastritis.  
CC The antibacterial agents are useful to treat in-dwelling devices for infection prevention or generally as wound treatments to prevent adhesion of bacteria to matrix proteins. The MucC polypeptide is also useful for diagnosing or prognosing a (susceptibility to) disease, for raising antibodies, to identify modulators or specific receptors, in rational drug design and as an immunogen for vaccines. The MucC gene sequences are useful in antisense/ribozyme therapeutics; to detect mutant MucC gene; for chromosomal mapping; to determine bacterial serotype; and for genetic immunisation.  
SQ Sequence 1351 BP; 471 A; 166 C; 260 G; 454 T;  
Query Match 100.0%; Score 19; DB 60; Length 1351;  
Best Local Similarity 100.0%; Pred. No. 9,94e-02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 865 gttaacaatattataagaag 883  
|||||  
QY 1 GTTACAAATATTAAAGAC 19  
RESULT 5  
ID V74703 standard; DNA; 2424 BP.  
AC V74703;  
DT 16-MAR-1999 (first entry)  
DE *Staphylococcus aureus* contig SEQ ID #392.  
KW Computer readable medium; vaccine; *S. aureus* infection; immunodetection; cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy; skin infection; surgical wound infection; scalded skin syndrome; ds.  
KW toxic shock syndrome; ds.  
KW *Staphylococcus aureus*.  
OS *Staphylococcus aureus*.  
FH Key location/Qualifiers  
FT misc\_feature 1141..1200  
FT /tag= "a  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"  
EP-786519-A2.  
PD 30-JUL-1997.  
PR 07-JAN-1997; 100117.  
PR 05-JAN-1996; US-009861.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
PI Rosen CA;  
PI WP1: 97-374922/35.  
DR Polynucleotide(s) and proteins derived from *Staphylococcus aureus*  
PT stored on computer readable medium and used in the production of  
PT anti-*S. aureus* vaccines  
PS Claim 1: Page 1287-1288; 3271pp: English.  
CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the *S. aureus* DNA sequences allows putative functions to be assigned so that protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are

```

CC likely to encode antigens have been identified and these polypeptides
CC are used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
CC
SO Sequence 2424 BP; 839 A; 302 C; 462 G; 759 T;

Query Match 100.0%; Score 19; DB 56; Length 2424;
Best Local Similarity 100.0%; Pred. No. 9,94e-02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1470 gttaacaattttaagaag 1488
Oy 1 GTTACAAATATTAAAGAG 19

RESULT 6
ID T22923 standard; cDNA to mRNA; 299 BP.
AC T22923;
DT 24-OCT-1996 (first entry)
DE Human gene signature HUMGS04648.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W0514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATSU/) MATSUBARA K.
PA (OKUBO/) OKUBO K.
PI Matsubara K, Okubo K;
PI WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1246; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
CC
SO Sequence 299 BP; 93 A; 48 C; 45 G; 102 T;

Query Match 84.2%; Score 16; DB 22; Length 299;
Best Local Similarity 100.0%; Pred. No. 3.98e+00;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 134 cttaaatatttgaac 149
Oy 16 CTTTAATATTGTAAC 1

RESULT 7
ID O60017 standard; cDNA; 439 BP.
AC O60017;

```



DE 16-MAR-1994 (first entry)  
 DE Human brain Expressed Sequence Tag EST01999.  
 KM Gene transcription product; genetic markers; tagging; in vivo;  
 KM transcription; mapping; locations; chromosomes; chromosomal; ss.  
 OS Homo sapiens.  
 PN W0916178-A.  
 PD 19-AUG-1993.  
 PF 12-FEB-1994; 001294.  
 PR 12-FEB-1994; US-837195.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 PI Adams MD, Moreno RF, Venter CJ.  
 DR WPI: 93-272882/34.  
 PT Enriched oligonucleotides and corresp. sequences - used as  
 PT markers for human genes transcribed in-vivo, facilitate tagging  
 PT of most human genes  
 PS Example 4; Page 268; 500PP; English.  
 CC The Expressed Sequence Tag was isolated from a human brain cDNA  
 CC library as part of a large set of ESTs which can be used as markers  
 CC for human genes transcribed in vivo. They can be used to facilitate  
 CC tagging of most human genes, for mapping locations of expressed genes  
 CC on chromosomes, for individual or forensic identification, for mapping  
 CC locations of disease-associated genes, for identification of tissue  
 CC type, and for prepn. of antisense sequences, probes and constructs.  
 CC EST1999 has a "marginal" coding probability as evaluated using the  
 CC coding-region prediction program CM. See also Q59041-Q61440.  
 SQ Sequence 439 BP; 101 A; 96 C; 81 G; 160 T;

Query Match 84.2%; Score 16; DB 8; Length 439;  
 Best Local Similarity 100.0%; Pred. No. 3,98e+00;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 188 ttcttaatatattgta 203  
 Cp 18 TTCTTAATATTGTA 3

RESULT 8  
 ID T75076 standard; DNA; 1767 BP.

AC T75076;

DE 11-SEP-1997 (first entry)

DE DNA encoding aspartyl-tRNA synthetase from Staph. aureus.

KM antibody; bacterial infection; adherence; damaged tissue; wound healing;

KM skin; protection; vaccine; ds.

OS Staphylococcus aureus.

PN EP-785265-A1.

PD 23-JUL-1997.

PF 17-JAN-1997; 300314.

PR 19-JAN-1996; GB-001096.

PR 30-OCT-1996; GB-022617.

PA (SMK ) SMITHKLINE BEECHAM PLC.

PI Hodgson JE, Lawlor EJ;

DR WPI: 97-365940/34.

DR P-PSDB: W22708.

PT DNA encoding aspartyl-tRNA synthetase from Staphylococcus aureus

PT WCUH 29 - useful for protection against bacterial infections

PS Claim 4; Page 18; 31pp; English.

CC The present sequence encodes a novel aspartyl tRNA synthetase protein

CC from Staphylococcus aureus strain WCUH29 (NCIMB 40771), which is related

CC by amino acid sequence homology to Bacillus subtilis tRNA synthetase

CC encoded by the tyrS gene. The enzyme catalyses the aminoacylation of a

CC stable enzyme:aspartyl adenylate complex resulting from the specific

CC binding and reaction of ATP and L-Aspartate. Subsequently, the 3'

CC terminal adenosine of enzyme-bound tRNA-asp reacts with the aminoacyl

CC adenylate, leading to the esterification of the tRNA and release of AMP.

CC Vectors comprising the DNA (or polynucleotides) having at least 70 %

CC identity to it) can be used for the recombinant production of the enzyme.

CC The enzyme or its related DNA (through gene therapy) is used to induce an

CC immunological response in a mammal to generate antibodies to protect

CC against disease. The antibodies protect against invasion of bacteria,

CC e.g. by blocking adherence of bacteria to damaged tissue, including

CC wounds in skin or connective tissue caused by mechanical, chemical or

CC thermal damage or by implantation of in-dwelling devices, or wounds in

CC the mucous membranes.

SQ Sequence 1767 BP; 614 A; 261 C; 367 G; 525 T;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 84.2%; Score 16; DB 31; Length 1767;  
 Best Local Similarity 100.0%; Pred. No. 3,98e+00;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 295 gttaacaataatgaag 310  
 QY 1 GTTACAAATATTGAAG 16

RESULT 9

ID V74671 standard; DNA; 2524 BP.

AC V74671;

DE 16-MAR-1999 (first entry)

DE Staphylococcus aureus confg SEQ ID #360.

KM Computer readable medium; vaccine; S.aureus infection; immunodetection;

KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

KM skin infection; surgical wound infection; scalded skin syndrome;

OS toxic shock syndrome; ds.

PN Staphylococcus aureus.

PR Key location/qualifiers

PR misc\_feature 841..900

FT /note= "these bases represent a line of missing text in

FT the sequence listing in the specification. They

FT are included to maintain the nucleotide numbering

FT given in the specification for this DNA sequence"

PN EP-786519-A2.

PD 30-JUL-1997.

PF 07-JAN-1997; 100117.

PR 05-JAN-1996; US-009861.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,

PI Rosen CA;

DR WPI: 97-374922/35.

PT Polynucleotide(s) and proteins derived from Staphylococcus aureus

PT stored on computer readable medium and used in the production of

PT anti-S.aureus vaccines

PS Claim 1; Page 1244-1245; 3271pp; English.

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences

CC of the invention. The DNA sequences are recorded on a computer readable

CC medium, preferably selected from a floppy or hard disk, random access

CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using

CC the S.aureus DNA sequences allows putative functions to be assigned so

CC that protein-encoding or regulatory regions of commercial, therapeutic or

CC industrial importance can be obtained. Specifically, sequences which are

CC likely to encode antigens have been identified and these polypeptides can

CC be used in a vaccine composition against S.aureus infection. The

CC polypeptides can also be used in a kit for the immunodetection of

CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,

CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,

CC skin and surgical wound infections, scalded skin syndrome, toxic shock

CC syndrome, etc. Organisms transformed with the DNA sequences can be used

CC for recombinant production of the polypeptides. The new DNA sequences

CC (and their fragments) are useful as primers or probes for isolating

CC homologues of any of the S.aureus DNA sequences contained on the

CC computer readable medium.

SQ Sequence 2524 BP; 876 A; 321 C; 517 G; 746 T;

Query Match 84.2%; Score 16; DB 58; Length 2524;  
 Best Local Similarity 100.0%; Pred. No. 3,98e+00;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1704 gttaacaataatgaag 1719  
 QY 1 GTTACAAATATTGAAG 16

RESULT 10  
 ID T40041 standard; DNA; 1003 BP.



AC T40041; (first entry)  
 DT 25-JUL-1997  
 DE Presentin-1 exon 12.  
 KW Presentin-1: human; hps1-2; PS-2; Integral membrane protein; AD; familial Alzheimer's disease; cerebral haemorrhage; schizophrenia; exon; depression; antibody; gene expression modulator; therapy; ss.  
 OS Homo sapiens.  
 PN W0634099-A2.  
 PD 31-OCT-1996.  
 PF 29-APR-1996; CA0263.  
 PR 28-APR-1995; US-431048.  
 PR 28-JUN-1995; US-496841.  
 PR 31-JUL-1995; US-509359.  
 PA (HSCR-) HSC RES & DEV LP.  
 PI (UTOR) UNIV TORONTO GOVERNING COUNCIL.  
 PR Fraser PE, Rommens JM, St George-Hyslop PH; WPI: 96-497631/49.  
 DR New presentin genes - useful for diagnosis, therapy and drug screening of familial Alzheimer's disease, cerebral disorders, etc.  
 PT Claim 8; Page 143-144; 178pp; English.  
 PS T40032-T40042 represent exons of the wild type human presentin-1 (PS-1) sequence (see T40028 and T40029 for the two cDNA forms of PS-1). The PS-1 form represented by T40029 results from alternate splicing of the genomic DNA sequence. T40031 represents the coding sequence for wild type human PS-2. The presentins are a family of highly conserved integral membrane proteins with a common structural motif, common alternate splicing patterns, and common mutational hot spot regions. Mutations in PS genes are implicated in familial Alzheimer's disease (AD) and possibly other diseases such as cerebral haemorrhage, schizophrenia, depression etc., so detection of mutations in these sequences can be used for diagnosis of these diseases. The encoded proteins, or vectors that express them or containing antisense sequences, antibodies selective for mutant forms of the encoded proteins (such as W05736) and modulators of PS gene expression are potentially useful for treatment of AD etc. Transgenic animals are useful as models for drug screening. The antibodies can also be used e.g. for affinity purification and in immunoassays.  
 CC Sequence 1003 BP; 292 A; 188 C; 193 G; 313 T;  
 SQ

Query Match 78.9%; Score 15; DB 30; Length 1003;  
 Best Local Similarity 100.0%; Pred. No. 1.31e+01;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 403 ttaataattgttaac 417  
 |||||  
 Cp 15 TTTAATATTGTGAC 1

RESULT 11  
 ID T99670 standard; DNA: 1003 BP.  
 AC T99670;  
 DT 20-JUL-1998 (first entry)  
 DE Human presentin-1 gene exon 12 and flanking regions.  
 KW Presentin-1; PS1 gene; human; familial Alzheimer's disease; FAD; cerebral haemorrhage; schizophrenia; depression; epilepsy; mental retardation; diagnosis; therapy; transgenic animal; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT exon 431..549  
 FT /\*tag= a  
 FT /number= 12

PN W09801549-A2.  
 PD 15-JAN-1998.  
 PF 04-JUL-1997; CA0475.  
 PR 02-JAN-1997; US-034590.  
 PR 05-JUL-1996; US-021673.  
 PR 12-JUL-1996; US-021700.  
 PR 08-NOV-1996; US-029895.  
 PA (HSCR-) HSC RES & DEV LP.  
 PI (UTOR) UNIV TORONTO GOVERNING COUNCIL.  
 PR Fraser PE, Rommens JM, St George-Hyslop PH; WPI: 96-286353/25.  
 DR New isolated mutant presentin-1 genes - useful for developing products for use in detection, diagnosis and therapy of Alzheimer's

FT disease and for drug screening  
 PS Disclosure; Page 195-196; 238pp; English.  
 CC This genomic DNA sequence comprises exon 12 plus flanking regions of the human presentin-1 (hps1) gene. Flanking genomic sequences (see T99661-71) for exons 1-13 of the hps1 gene, as well as hps1 cDNA clones (see V04666-67), are provided. Mutations in the human presentin genes have been linked to the development of familial Alzheimer's disease (FAD). Presentin nucleic acids and proteins can be used in screening and diagnosing FAD, identifying CC and developing therapeutics for treatment of FAD, and in producing cell lines and transgenic animals useful as models of FAD.  
 CC Sequence 1003 BP; 292 A; 188 C; 193 G; 313 T;  
 SQ

Query Match 78.9%; Score 15; DB 41; Length 1003;  
 Best Local Similarity 100.0%; Pred. No. 1.31e+01;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 403 ttaataattgttaac 417  
 |||||  
 Cp 15 TTTAATATTGTGAC 1

RESULT 12  
 ID T73866 standard; DNA: 3045 BP.  
 AC T73866;  
 DT 26-JAN-1998 (first entry)  
 DE Cotton fibre promoter clone Rac13 construct, PCGN4735.  
 KW Promoter; fibre-specific; transcriptional factor; promoter; altered phenotype; colour; melanin; indigo; ss.  
 OS Gossypium hirsutum cv. Coker 130.  
 PN W09640924-A2.  
 PD 19-DEC-1996.  
 PF 07-JUN-1996; U09897.  
 PR 07-JUN-1995; US-480178.  
 PR 01-JUL-1996; ZA-005572.  
 PA (CALJ) CALGENE INC.  
 PI McBride K, Pear JR, Perez-Grau L, Stalker DM; WPI: 97-052325/05.  
 DR DNA construct contg. gene of interest controlled by cotton fibre PT transcriptional factor - used to produce altered phenotype cotton PT fibre cells expressing genes affecting pigmentation  
 PS Claim 23; Fig 5A-E; 95pp; English.  
 CC The present sequence is the Rac13 promoter construct, PCGN4735, isolated from cotton fibre genomic clone 15-1. DNA constructs containing CC cotton fibre-specific transcriptional factor promoters are useful to CC produce cotton fibre cells with altered phenotype, especially altered colour. Genes involved in the production of melanin (e.g. tyrosinase CC gene and ORF438 encoded protein from Streptomyces antibioticus) and CC indigo (mono-oxygenase genes possibly in conjunction with a CC tryptophanase gene) are of interest. The promoters of the invention are CC reliable and permit expression of a protein selectively in cotton fibre CC to affect qualities such as fibre strength, length, colour and dyability CC as required. The construct and methods can also be used for the CC introduction of other advantageous genes into a cotton plant, e.g. a CC plant hormone. In particular, fibres from a plant producing coloured CC fibres may be used to produce yarns and/or fabrics that do not require CC dyeing.  
 CC Sequence 3045 BP; 1063 A; 450 C; 366 G; 1162 T;  
 SQ

Query Match 78.9%; Score 15; DB 34; Length 3045;  
 Best Local Similarity 100.0%; Pred. No. 1.31e+01;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 192 tcttaataattgtta 206  
 |||||  
 Cp 17 TCTTAATATTGTA 3

RESULT 13  
 ID V74407 standard; DNA: 11050 BP.  
 AC V74407;  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SEQ ID #96.



KW Computer readable medium; vaccine; *S.aureus* infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.  
 OS *Staphylococcus aureus*.  
 FH key  
 FT misc-feature  
 FT Location/Qualifiers  
 FT 781..840  
 FT /tag- a  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT 2581..2640  
 FT /tag- b  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT 4381..4440  
 FT /tag- c  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT 6181..6240  
 FT /tag- d  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT 7981..8040  
 FT /tag- e  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature  
 FT 9781..9840  
 FT /tag- f  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 PN EP-786519-A2.  
 PD 30-JUL-1997.  
 PF 07-JAN-1997: 100117.  
 PR 05-JAN-1996: US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
 PI Rosen CA,  
 DR WPI: 97-374922/35.  
 PT Polynucleotide(s) and proteins derived from *Staphylococcus aureus* -  
 PT stored on computer readable medium and used in the production of  
 PT anti-*S.aureus* vaccines  
 PS Claim 1: Page 577-583: 3271pp. English.  
 CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences  
 of the invention. The DNA sequences are recorded on a computer readable  
 medium, preferably selected from a floppy or hard disk, random access  
 memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 the *S.aureus* DNA sequences allows putative functions to be assigned so  
 that protein-encoding or regulatory regions of commercial, therapeutic or  
 industrial importance can be obtained. Specifically, sequences which are  
 likely to encode antigens have been identified and these polypeptides can  
 be used in a vaccine composition against *S.aureus* infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC *S.aureus* in a sample. *S.aureus* is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the *S.aureus* DNA sequences contained on the  
 CC computer readable medium.  
 SQ Sequence 11050 BP: 3763 A: 1649 C: 1897 G: 3373 T:

Query Match 78.9%; Score 15; DB 58; Length 11050;  
 Best Local Similarity 100.0%; Pred. No. 1.31e+01;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 8069 ttacaattatgaag 8083  
 Qy 2 TTACCAATATTGAAG 16  
 ID 119814 standard: cDNA to mRNA; 82 BP.  
 AC T19814;  
 DT 12-JUL-1996 (first entry)  
 DE Human gene signature HUMGS00892.  
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;  
 KW cell typing; abnormal cell function; ss.  
 OS Homo sapiens.  
 PN W09514772-A1.  
 PD 01-JUN-1995.  
 PF 11-NOV-1994: J01916.  
 PR 12-NOV-1993: JP-335504.  
 PA (MATS/) MATSUBARA K.  
 PI (OKUB/) OKUBO K.  
 PI Matsubara K, Okubo K;  
 DR WPI: 95-206931/27.  
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing cDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues  
 PS Claim 1: Page 477; 2245pp; Japanese.  
 CC A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in T19001-T26837 and which is able to hybridise to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared  
 CC from various human tissues; synthesis of cDNA was initiated from the  
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
 CC is constructed so as to reflect accurately the relative abundance of  
 CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.  
 SQ Sequence 82 BP: 35 A: 9 C: 6 G: 31 T;  
 Query Match 73.7%; Score 14; DB 19; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 4.20e+01;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 68 tttaattttgtaa 81  
 Cp 15 TTTAATATTGTA 2  
 ID V89012 standard: cDNA; 250 BP.  
 AC V89012;  
 DT 12-FEB-1999 (first entry)  
 DE EST clone 17942.  
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolytics;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
 OS Homo sapiens.  
 PN W09845437-A2.  
 PD 15-OCT-1998.  
 PF 10-APR-1998: U06956.  
 PR 10-APR-1997: US-837312.  
 PA (GEMV ) GENETICS INST INC.



PI Agostino MJ, Jacobs K, Lavalie ER, McCoy JM, Werberg D,  
PI Racie LA, Spaulding V, Treacy M;  
DR WPI: 99-070078/06.  
PI New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries  
PS Claim 1, Page 588; 641pp; English.  
CC The present sequence represents an expressed sequence tag (EST), and is  
CC a polynucleotide of the invention. The polynucleotides of the invention  
CC are all secreted EST sequences isolated from a variety of human tissue  
CC sources. The EST sequences and proteins encoded by them are predicted to  
CC have useful biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and  
CC animals, although no supporting data is given. Suggested activities  
CC include nutritional activity, immune stimulating or suppressing activity,  
CC haematopoiesis regulating activity, tissue growth activity,  
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The EST sequences are also stated to be useful for gene  
CC therapy.  
SQ Sequence 250 BP; 69 A; 41 C; 58 G; 82 T;

Query Match 73.7%; Score 14; DB 53; Length 250;  
Best Local Similarity 100.0%; Pred. No. 4.20e+01;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

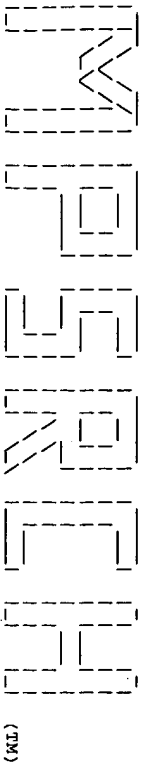
DB 68 tcttaatttgt 81  
CP 17 tctttaatttgt 4

Search completed: Sat Nov 27 15:25:08 1999  
Job time : 55 secs.



THIRTY EIGHT  
BLANK (USPTO)





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*

Run on: Sat Nov 27 15:25:54 1999; MasPar time 4.50 Seconds  
365.426 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-6  
Description: (1-19) from US09103287.seq  
Perfect Score: 19  
N.A. Sequence: 1 GTTACAATATTAAAGAG 19  
Comp: CATGTTTAAATTCTTC

Scoring table: TABLE jmetric  
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 165359 segs, 43243793 bases x 2

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: n-issued  
1:5A\_COMB 2:5B\_COMB 3:5C\_COMB 4:PC19\_COMB 5:backfiles1

Statistics: Mean 4.284; Variance 2.297; scale 1.865

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	15	78.9	1003	3	US-08-967- Sequence 151, Applicat	8.09e+01
2	14	73.7	2029	5	5495001-6 Patent No. 5495001.	3.30e+00
3	14	73.7	2029	5	5457090-1 Patent No. 5457090.	3.30e+00
4	14	73.7	2029	5	5187089-1 Patent No. 5187089.	3.30e+00
5	14	73.7	2031	5	5495001-8 Patent No. 5495001.	3.30e+00
6	14	73.7	2032	5	5457090-3 Patent No. 5457090.	3.30e+00
7	14	73.7	2032	5	5187089-3 Patent No. 5187089.	3.30e+00
8	14	73.7	8201	2	US-08-253- Sequence 9, Applicatio	1.30e+01
9	13	68.4	345	1	US-08-393- Sequence 12, Applicati	1.30e+01
10	13	68.4	345	1	US-08-393- Sequence 12, Applicati	1.30e+01
11	13	68.4	345	2	US-08-726- Sequence 12, Applicati	1.30e+01
12	13	68.4	345	2	US-08-726- Sequence 12, Applicati	1.30e+01
13	13	68.4	489	4	US-08-726- Sequence 13, Applicati	1.30e+01
14	13	68.4	489	4	PCT-US94-0 Sequence 4, Applicatio	1.30e+01
15	13	68.4	489	4	US-08-535- Sequence 4, Applicatio	1.30e+01
16	13	68.4	489	1	US-08-393- Sequence 1, Applicatio	1.30e+01
17	13	68.4	489	3	US-08-392- Sequence 4, Applicatio	1.30e+01
18	13	68.4	489	4	PCT-US96-0 Sequence 1, Applicatio	1.30e+01
19	13	68.4	489	4	US-08-393- Sequence 1, Applicatio	1.30e+01
20	13	68.4	489	2	US-08-504- Sequence 1, Applicatio	1.30e+01

21	13	68.4	489	2	US-08-504- Sequence 4, Applicatio	1.30e+01
22	13	68.4	489	1	US-08-031- Sequence 1, Applicatio	1.30e+01
23	13	68.4	489	1	US-08-031- Sequence 4, Applicatio	1.30e+01
24	13	68.4	489	2	US-08-726- Sequence 1, Applicatio	1.30e+01
25	13	68.4	489	4	PCT-US94-0 Sequence 1, Applicatio	1.30e+01
26	13	68.4	489	3	US-08-392- Sequence 2, Applicatio	1.30e+01
27	13	68.4	1198	3	US-08-396- Sequence 3, Applicatio	1.30e+01
28	13	68.4	1198	1	US-08-020- Sequence 3, Applicatio	1.30e+01
29	13	68.4	1469	1	US-08-488- Sequence 8, Applicatio	1.30e+01
30	13	68.4	1469	4	PCT-US96-0 Sequence 8, Applicatio	1.30e+01
31	13	68.4	1469	4	US-08-820- Sequence 41, Applicati	1.30e+01
32	13	68.4	1469	1	US-08-488- Sequence 2, Applicatio	1.30e+01
33	13	68.4	1745	1	PCT-US96-0 Sequence 2, Applicatio	1.30e+01
34	13	68.4	1745	4	PCT-US96-0 Sequence 13, Applicati	1.30e+01
35	13	68.4	1950	3	US-08-472- Sequence 30, Applicati	1.30e+01
36	13	68.4	2015	3	US-08-417- Sequence 7, Applicatio	1.30e+01
37	13	68.4	2020	3	US-08-417- Sequence 136, Applicat	1.30e+01
38	13	68.4	2028	3	US-08-417- Sequence 139, Applicat	1.30e+01
39	13	68.4	2057	3	US-08-820- Sequence 42, Applicati	1.30e+01
40	13	68.4	2060	3	US-08-417- Sequence 142, Applicati	1.30e+01
41	13	68.4	2233	3	US-08-129- Sequence 2, Applicatio	1.30e+01
42	13	68.4	2304	1	US-08-020- Sequence 1, Applicatio	1.30e+01
43	13	68.4	2304	3	US-08-396- Sequence 1, Applicatio	1.30e+01
44	13	68.4	2304	3	US-08-417- Sequence 79, Applicati	1.30e+01
45	13	68.4	5496	1	US-08-181- Sequence 2, Applicatio	1.30e+01

# ALIGNMENTS

RESULT 1  
ID US-08-967-101-151 STANDARD; DNA; UNC; 1003 BP.  
AC xxxxxx  
DT  
DE Sequence 151, Application US/08967101  
CC Patent No. 5840540  
CC GENERAL INFORMATION:  
CC APPLICANT: ST. GEORGE-HYSLOP, PETER H  
CC APPLICANT: ROWMENS, JOHANNA M  
CC APPLICANT: FRASER, PAUL E  
CC TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
CC NUMBER OF SEQUENCES: 183  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
CC STREET: High Street Tower - 125 High Street  
CC CITY: Boston  
CC STATE: Massachusetts  
CC COUNTRY: U.S.A.  
CC ZIP: 02110  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/967,101  
CC FILING DATE: 10-NOV-1997  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/592,541  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Fletcher, Edmund R.  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 248-7100  
CC TELEFAX: (617) 248-7100  
CC INFORMATION FOR SEQ ID NO: 151:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1003 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear



CC MOLECULE TYPE: CDNA  
SQ SEQUENCE 1003 BP: 292 A; 188 C; 193 G; 313 T; 17 OTHER.

Query Match 78.9%; Score 15; DB 3; Length 1003;  
Best Local Similarity 100.0%; Pred. No. 8.09e-01;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 403 TTTAATATTGTAC 417  
15 TTTAATATTGTAC 1

RESULT 2  
ID 5495001-6 STANDARD; DNA; UNC; 2198 BP.  
AC xxxxxx  
DT 01-JAN-1900

DE Patent No. 5495001.

CC APPLICANT: MCGROGAN, MICHAEL P.; SCOTT, RANDY W.; BAKER,  
CC JOFFRE B.; SIMONSEN, CHRISTIAN C.

CC TITLE OF INVENTION: RECOMBINANT PURIFIED PROTEASE NEXIN  
CC NUMBER OF SEQUENCES: 16  
CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/288,596  
CC FILING DATE: 10-AUG-1994  
CC PRIORITY APPLICATION DATA:

CC APPLICATION NUMBER: 111,111  
CC FILING DATE: 24-AUG-1993  
CC APPLICATION NUMBER: 25,450

CC FILING DATE: 13-MAR-1987  
CC APPLICATION NUMBER: 871,501  
CC FILING DATE: 06-JUN-1986

CC APPLICATION NUMBER: 870,232  
CC FILING DATE: 03-JUN-1986  
CC SEQ ID NO: 6;

CC LENGTH: 2029  
SQ Sequence 2198 BP: 536 A; 477 C; 451 G; 565 T; 169 other;

Query Match 73.7%; Score 14; DB 5; Length 2029;  
Best Local Similarity 100.0%; Pred. No. 3.30e+00;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 413 AAATATTAAAGAG 426  
6 AAATATTAAAGAG 19

RESULT 3  
ID 5457090-1 STANDARD; DNA; UNC; 2198 BP.  
AC xxxxxx  
DT 01-JAN-1900

DE Patent No. 5457090.  
CC Patent No. 5457090

CC APPLICANT: SCOTT, RANDY W.; GOLINI, FRED; MCGROGAN, MICHAEL  
CC TITLE OF INVENTION: PROTEASE NEXIN-I VARIANTS  
CC NUMBER OF SEQUENCES: 7  
CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/07/924,294  
CC FILING DATE: 03-AUG-1992  
CC PRIORITY APPLICATION DATA:

CC APPLICATION NUMBER: 542,484  
CC FILING DATE: 21-JUN-1990  
CC SEQ ID NO: 1;

CC LENGTH: 2029  
SQ Sequence 2198 BP: 534 A; 476 C; 455 G; 564 T; 169 other;

Query Match 73.7%; Score 14; DB 5; Length 2029;  
Best Local Similarity 100.0%; Pred. No. 3.30e+00;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 413 AAATATTAAAGAG 426  
6 AAATATTAAAGAG 19

QY 6 AAATATTAAAGAG 19

RESULT 4  
ID 5187089-1 STANDARD; DNA; UNC; 2198 BP.  
AC xxxxxx  
DT 01-JAN-1900

DE Patent No. 5187089.  
CC Patent No. 5187089

CC APPLICANT: SCOTT, RANDY W.; GOLINI, FRED; MCGROGAN, MICHAEL  
CC TITLE OF INVENTION: PROTEASE NEXIN-I VARIANTS WHICH INHIBIT  
CC ELASTASE  
CC NUMBER OF SEQUENCES: 11  
CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/07/542,484  
CC FILING DATE: 21-JUN-1990  
CC SEQ ID NO: 1;

CC LENGTH: 2029  
SQ Sequence 2198 BP: 534 A; 476 C; 455 G; 564 T; 169 other;

Query Match 73.7%; Score 14; DB 5; Length 2029;  
Best Local Similarity 100.0%; Pred. No. 3.30e+00;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 413 AAATATTAAAGAG 426  
6 AAATATTAAAGAG 19

RESULT 5  
ID 5495001-8 STANDARD; DNA; UNC; 2201 BP.  
AC xxxxxx  
DT 01-JAN-1900

DE Patent No. 5495001.  
CC Patent No. 5495001

CC APPLICANT: MCGROGAN, MICHAEL P.; SCOTT, RANDY W.; BAKER,  
CC JOFFRE B.; SIMONSEN, CHRISTIAN C.

CC TITLE OF INVENTION: RECOMBINANT PURIFIED PROTEASE NEXIN  
CC NUMBER OF SEQUENCES: 16  
CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/288,596  
CC FILING DATE: 10-AUG-1994  
CC PRIORITY APPLICATION DATA:

CC APPLICATION NUMBER: 111,111  
CC FILING DATE: 24-AUG-1993  
CC APPLICATION NUMBER: 25,450

CC FILING DATE: 13-MAR-1987  
CC APPLICATION NUMBER: 871,501  
CC FILING DATE: 06-JUN-1986

CC APPLICATION NUMBER: 870,232  
CC FILING DATE: 03-JUN-1986  
CC SEQ ID NO: 8;

CC LENGTH: 2031  
SQ Sequence 2201 BP: 537 A; 480 C; 449 G; 565 T; 170 other;

Query Match 73.7%; Score 14; DB 5; Length 2031;  
Best Local Similarity 100.0%; Pred. No. 3.30e+00;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 412 AAATATTAAAGAG 425  
6 AAATATTAAAGAG 19

RESULT 6  
ID 5457090-3 STANDARD; DNA; UNC; 2202 BP.  
AC xxxxxx  
DT 01-JAN-1900

DE Patent No. 5457090.  
CC Patent No. 5457090

CC APPLICANT: SCOTT, RANDY W.; GOLINI, FRED; MCGROGAN, MICHAEL  
CC TITLE OF INVENTION: PROTEASE NEXIN-I VARIANTS  
CC NUMBER OF SEQUENCES: 7  
CC CURRENT APPLICATION DATA:

QY 6 AAATATTAAAGAG 19



CC APPLICATION NUMBER: US/07/924,294  
 CC FILING DATE: 03-AUG-1992  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 542,484  
 CC FILING DATE: 21-JUN-1990  
 CC SEQ ID NO:3  
 CC LENGTH: 2032  
 CC Sequence 2202 BP; 536 A; 479 C; 454 G; 563 T; 170 other;  
 Query Match 73.7%; Score 14; DB 5; Length 2032;  
 Best Local Similarity 100.0%; Pred. No. 3.30e+00;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 413 AATATTAAAGAG 426  
 QY 6 AATATTAAAGAG 19  
 RESULT 7  
 ID 5187089-3 STANDARD; DNA; UNC; 2202 BP.  
 AC xxxxxx  
 DT 01-JAN-1900  
 DE Patent No. 5187089.  
 CC Patent No. 5187089  
 CC APPLICANT: SCOTT, RANDY W.; GOLINI, FRED; MCGROGAN, MICHAEL  
 CC TITLE OF INVENTION: PROTEASE NEXIN-I VARIANTS WHICH INHIBIT  
 CC ELASTASE  
 CC NUMBER OF SEQUENCES: 11  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/07/542,484  
 CC FILING DATE: 21-JUN-1990  
 CC SEQ ID NO:3  
 CC LENGTH: 2032  
 CC Sequence 2202 BP; 535 A; 477 C; 456 G; 564 T; 170 other;  
 Query Match 73.7%; Score 14; DB 5; Length 2032;  
 Best Local Similarity 100.0%; Pred. No. 3.30e+00;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 413 AATATTAAAGAG 426  
 QY 6 AATATTAAAGAG 19  
 RESULT 8  
 ID US-08-253-155A-9 STANDARD; DNA; UNC; 8201 BP.  
 AC xxxxxx  
 DE Sequence 9, Application US/08253155A  
 CC Sequence 9, Application US/08253155A  
 CC Patent No. 5691147  
 CC GENERAL INFORMATION:  
 CC APPLICANT: GYURIS, JENO  
 CC APPLICANT: Diacteta, Giulio  
 CC TITLE OF INVENTION: CDK4 Binding Proteins  
 CC NUMBER OF SEQUENCES: 95  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: LAHYE & COCKFIELD  
 CC STREET: 60 State Street  
 CC CITY: Boston  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02109  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: ASCII(text)  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/253,155A  
 CC FILING DATE: 02-JUN-1994  
 CC CLASSIFICATION: 435  
 CC ATTORNEY/AGENT INFORMATION:

CC NAME: Vincent, Matthew P.  
 CC REGISTRATION NUMBER: 36,709  
 CC REFERENCE/DOCKET NUMBER: MIT-028  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 9:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 8201 base pairs  
 CC TYPE: nucleic acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: CDNA  
 CC SEQUENCE 8201 BP; 2258 A; 1887 C; 2003 G; 2053 T; 0 OTHER.  
 Query Match 73.7%; Score 14; DB 2; Length 8201;  
 Best Local Similarity 100.0%; Pred. No. 3.30e+00;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 6464 ACAATATTAAAGA 6477  
 QY 4 ACAATATTAAAGA 17  
 RESULT 9  
 ID US-08-393-305-12 STANDARD; DNA; UNC; 345 BP.  
 AC xxxxxx  
 DE Sequence 12, Application US/08393305  
 CC Sequence 12, Application US/08393305  
 CC Patent No. 5574138  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Grabsteijn, Kenneth  
 CC APPLICANT: Anderson, Dirk  
 CC APPLICANT: Eisenman, June  
 CC APPLICANT: Fung, Victor  
 CC APPLICANT: Rauch, Charles  
 CC TITLE OF INVENTION: EPITHELIAL-DERIVED T-CELL FACTOR  
 CC NUMBER OF SEQUENCES: 15  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Seed and Berry  
 CC STREET: 6300 Columbia Center, 701 Fifth Avenue  
 CC CITY: Seattle  
 CC STATE: Washington  
 CC COUNTRY: USA  
 CC ZIP: 98104  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/393,305  
 CC FILING DATE: 22-FEB-1995  
 CC CLASSIFICATION: 424  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Mcmasters, David D.  
 CC REGISTRATION NUMBER: 33,963  
 CC REFERENCE/DOCKET NUMBER: 480052.409C2  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: 206-622-4900  
 CC INFORMATION FOR SEQ ID NO: 12:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 345 base pairs  
 CC TYPE: nucleic acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC FEATURE:  
 CC NAME/KEY: CDS  
 CC LOCATION: 1..345  
 CC SEQUENCE 345 BP; 121 A; 49 C; 68 G; 107 T; 0 OTHER.  
 Query Match 68.4%; Score 13; DB 1; Length 345;



Best Local Similarity 100.0%; Pred. No. 1.30e+01;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 282 AAATATTAAAGAA 294  
QY 6 AAATATTAAAGAA 18

RESULT 10  
ID US-08-393-305-13 STANDARD; DNA; UNC; 345 BP.  
AC xxxxxx

DE Sequence 13, Application US/08393305  
CC Sequence 13, Application US/08393305  
CC Patent No. 5574138

CC GENERAL INFORMATION:

CC APPLICANT: Grabstein, Kenneth

CC APPLICANT: Anderson, Dirk

CC APPLICANT: Eisenman, June

CC APPLICANT: Fung, Victor

CC APPLICANT: Rauch, Charles

CC TITLE OF INVENTION: EPITHELIAL-DERIVED T-CELL FACTOR

CC NUMBER OF SEQUENCES: 15

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Seed and Berry

CC STREET: 6300 Columbia Center, 701 Fifth Avenue

CC CITY: Seattle

CC STATE: Washington

CC COUNTRY: USA

CC ZIP: 98104

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/393,305

CC FILING DATE: 22-FEB-1995

CC CLASSIFICATION: 424

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Mcmasters, David D.

CC REGISTRATION NUMBER: 33,963

CC REFERENCE/DOCKET NUMBER: 480052.409C2

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 206-622-4900

CC INFORMATION FOR SEQ ID NO: 13:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 345 base pairs

CC TYPE: nucleic acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC FEATURE:

CC NAME/KEY: CDS

CC LOCATION: 1..345

CC SEQUENCE 345 BP; 124 A; 50 C; 64 G; 107 T; 0 OTHER.

CC Query Match 68.4%; Score 13; DB 1; Length 345;

CC Best Local Similarity 100.0%; Pred. No. 1.30e+01;

CC Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 282 AAATATTAAAGAA 294

QY 6 AAATATTAAAGAA 18

RESULT 11

ID US-08-726-817-12 STANDARD; DNA; UNC; 345 BP.

AC xxxxxx

DE Sequence 12, Application US/08726817

CC Sequence 12, Application US/08726817

CC Patent No. 5707616

CC GENERAL INFORMATION:

CC APPLICANT: Grabstein, Kenneth  
CC APPLICANT: Anderson, Dirk  
CC APPLICANT: Eisenman, June  
CC APPLICANT: Fung, Victor  
CC APPLICANT: Rauch, Charles  
CC TITLE OF INVENTION: EPITHELIAL-DERIVED T-CELL FACTOR

CC NUMBER OF SEQUENCES: 15

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Seed and Berry

CC STREET: 6300 Columbia Center, 701 Fifth Avenue

CC CITY: Seattle

CC STATE: Washington

CC COUNTRY: USA

CC ZIP: 98104

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/726,817

CC FILING DATE: 04-OCT-1996

CC CLASSIFICATION: 435

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/393,305

CC FILING DATE: 22-FEB-1995

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Mcmasters, David D.

CC REGISTRATION NUMBER: 33,963

CC REFERENCE/DOCKET NUMBER: 480052.409C2

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 206-622-4900

CC INFORMATION FOR SEQ ID NO: 12:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 345 base pairs

CC TYPE: nucleic acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC FEATURE:

CC NAME/KEY: CDS

CC LOCATION: 1..345

CC SEQUENCE 345 BP; 121 A; 49 C; 68 G; 107 T; 0 OTHER.

CC Query Match 68.4%; Score 13; DB 2; Length 345;

CC Best Local Similarity 100.0%; Pred. No. 1.30e+01;

CC Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 282 AAATATTAAAGAA 294

QY 6 AAATATTAAAGAA 18

RESULT 12

ID US-08-726-817-13 STANDARD; DNA; UNC; 345 BP.

AC xxxxxx

DE Sequence 13, Application US/08726817

CC Sequence 13, Application US/08726817

CC Patent No. 5707616

CC GENERAL INFORMATION:

CC APPLICANT: Grabstein, Kenneth

CC APPLICANT: Anderson, Dirk

CC APPLICANT: Eisenman, June

CC APPLICANT: Fung, Victor

CC APPLICANT: Rauch, Charles

CC TITLE OF INVENTION: EPITHELIAL-DERIVED T-CELL FACTOR

CC NUMBER OF SEQUENCES: 15

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Seed and Berry

CC STREET: 6300 Columbia Center, 701 Fifth Avenue

CC CITY: Seattle

CC STATE: Washington

CC COUNTRY: USA



```

CC ZIP: 98104
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/726,817
CC FILING DATE: 04-OCT-1996
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/393,305
CC FILING DATE: 22-FEB-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Mcmasters, David D.
CC REGISTRATION NUMBER: 33,963
CC REFERENCE/DOCKET NUMBER: 480052.409C2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 206-622-4900
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 345 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC FEATURE:
CC NAME/KEY: CDS
CC LOCATION: 1..345
CC SEQUENCE 345 BP; 124 A; 50 C; 64 G; 107 T; 0 OTHER.

```

	68.4%;	Score 13;	DB 2;	Length 345;
Query Match	Best Local Similarity 100.0%;	Pred. No.	1.30e+01;	
Matches	13; Conservative	Mismatches	0;	Indels 0; Gaps 0;
Dd	282 AATATTATAAGCA	294		
Oy	6 AATATTATAAGCA	18		

RESULT 13  
 ID US-08-726-817-4 STANDARD; DNA; UNC; 489 BP.  
 AC xxxxxx  
 DT  
 DE Sequence 4, Application US/08726817  
 CC Sequence 4, Application US/08726817  
 CC Patent No. 5707616  
 CC  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Grabstein, Kenneth  
 CC APPLICANT: Anderson, Dirk  
 CC APPLICANT: Eisenman, June  
 CC APPLICANT: Fung, Victor  
 CC APPLICANT: Rauch, Charles  
 CC TITLE OF INVENTION: EPITHELIUM-DERIVED T-CELL FACTOR  
 CC NUMBER OF SEQUENCES: 15  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Seed and Berry  
 CC STREET: 6300 Columbia Center, 701 Fifth Avenue  
 CC CITY: Seattle  
 CC STATE: Washington  
 CC COUNTRY: USA  
 CC ZIP: 98104  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/726,817  
 CC FILING DATE: 04-OCT-1996  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/393,305  
 CC FILING DATE: 22-FEB-1995  
 CC ATTORNEY/AGENT INFORMATION:  
 CC

CC NAME: Mcmasters, David D.  
CC REGISTRATION NUMBER: 33,963  
CC REFERENCE/DOCKET NUMBER: 480052.409C2  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 206-622-4900  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 489 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 1..489  
CC SEQUENCE 489 BP; 162 A; 81 C; 92 G; 154 T; 0 OTHER.

Query Match	68.4%	Score 13	DB 2	Length 489
Best Local Similarity	100.0%	Pred. No.	1.30e+01	
Matches	13	Conservative	0	Mismatches 0
				Indels 0
				Gaps 0
Db	426	AAATATTAAAGAA	438	
Oy	6	AAATATTAAAGAA	18	

```

DE      4, Application PC/US9403793
CC      Sequence 4, Application PC/US9403793
CC      GENERAL INFORMATION:
CC      APPLICANT: Grabstein, Kenneth
CC      APPLICANT: Anderson, Dirk
CC      APPLICANT: Eisenman, June
CC      APPLICANT: Fung, Victor
CC      APPLICANT: Rauch, Charles
CC      TITLE OF INVENTION: Interleukin-15
CC      NUMBER OF SEQUENCES: 12
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Immunex Corporation
CC      STREET: 51 University Street
CC      CITY: Seattle
CC      STATE: Washington
CC      COUNTRY: USA
CC      ZIP: 98101
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: Patent-In Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: PCT/US94/03793
CC      FILING DATE:
CC      CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Lauer, Charlene
CC      REGISTRATION NUMBER: 33,035
CC      REFERENCE/DOCKET NUMBER: 2811
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: 206-587-0430
CC      INFORMATION FOR SEQ ID NO: 4:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 489 base pairs
CC      TYPE: nucleic acid
CC      STRANDEDNESS: single
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: cDNA
CC      FEATURE:
CC      NAME/KEY: CDS
CC      LOCATION: 1..489
SQ      SEQUENCE 489 BP: 159 A; 79 C; 95 G; 156 T; 0 OTHER.

```



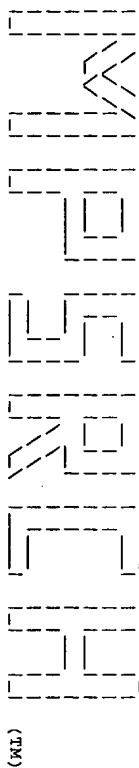
Query Match 68.4%; Score 13; DB 4; Length 489;  
 Best Local Similarity 100.0%; Pred. No. 1.30e+01;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 426 AATATTAAAGAA 438  
 QY 6 AATATTAAAGAA 18

RESULT 15  
 ID US-08-535-733-1 STANDARD; DNA; UNC; 489 BP.  
 AC xxxxxx  
 DT Sequence 1, Application US/08535733  
 DE Sequence 1, Application US/08535733  
 CC Patent No. 5660824  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Quinn, LeBris  
 CC APPLICANT: Grabstein, Kenneth  
 CC APPLICANT: Trout, Anthony B.  
 CC TITLE OF INVENTION: Muscle-Trophic Factor  
 CC NUMBER OF SEQUENCES: 2  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunex Corporation  
 CC STREET: 51 University Street  
 CC CITY: Seattle  
 CC STATE: Washington  
 CC COUNTRY: USA  
 CC ZIP: 98101  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: Apple Macintosh  
 CC OPERATING SYSTEM: System 7, Word 5.1a  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/535,733  
 CC FILING DATE:  
 CC CLASSIFICATION: 424  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Malaska, Stephen L.  
 CC REGISTRATION NUMBER: 32,655  
 CC REFERENCE/DOCKET NUMBER: 2833  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: 206-587-0430  
 CC INFORMATION FOR SEQ ID NO: 1:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 489 base pairs  
 CC TYPE: nucleic acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: CDNA  
 CC FEATURE:  
 CC NAME/KEY: CDS  
 CC LOCATION: 1..489  
 CC SEQUENCE 489 BP; 159 A; 79 C; 95 G; 156 T; 0 OTHER.

Query Match 68.4%; Score 13; DB 2; Length 489;  
 Best Local Similarity 100.0%; Pred. No. 1.30e+01;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 426 AATATTAAAGAA 438  
 QY 6 AATATTAAAGAA 18

Search completed: Sat Nov 27 15:26:08 1999  
 Job time : 14 secs.





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*

Run on: Sat Nov 27 15:20:04 1999; Maspar time 92.66 Seconds  
480.474 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-6  
Description: (1-19) from US09103287.seq  
Perfect Score: 19  
N.A. Sequence: 1 GTTACAAATATTAAGAG 19  
Comp: CAATGTTATATTTCTTC

Scoring table: TABLE Jmetric  
Gap 60

Mismatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb1-est58  
1:em\_est10.2:em\_est11.3:em\_est17.4:em\_est18.5:em\_est2  
6:em\_est9.7:em\_gss1

Database: genbank-est11  
8:gb\_est1.9:gb\_est10.10:gb\_est11.11:gb\_est12.12:gb\_est13  
13:gb\_est14.14:gb\_est15.15:gb\_est16.16:gb\_est17  
17:gb\_est18.18:gb\_est19.19:gb\_est20.20:gb\_est21  
21:gb\_est22.22:gb\_est23.23:gb\_est24.24:gb\_est25  
25:gb\_est26.26:gb\_est27.27:gb\_est28.28:gb\_est29  
29:gb\_est30.30:gb\_est31.31:gb\_est32.32:gb\_est33.33:gb\_est34  
34:gb\_est35.35:gb\_est36.36:gb\_est37.37:gb\_gss1.38:gb\_gss2  
39:gb\_gss3.40:gb\_gss4.41:gb\_gss5.42:gb\_gss6

Statistics: Mean 6.213; Variance 1.402; scale 4.431

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	18	94.7	551	15	AA583407	2.43e-05
2	17	89.5	299	25	A0037905	4.61e-04
3	16	84.2	253	8	Z39615	8.02e-03
4	16	84.2	298	36	C20646	8.02e-03
5	16	84.2	324	26	A1383567	8.02e-03
6	16	84.2	330	27	A1474210	8.02e-03
7	16	84.2	346	8	T05047	8.02e-03
8	16	84.2	350	21	AA988210	8.02e-03
9	16	84.2	369	21	AA936264	8.02e-03
10	16	84.2	380	36	AA122490	8.02e-03

11	16	84.2	409	24	A1214589	8.02e-03
12	16	84.2	410	26	A1365619	8.02e-03
13	16	84.2	414	12	AA349916	8.02e-03
14	16	84.2	415	19	R07577	8.02e-03
15	16	84.2	424	26	AA900060	8.02e-03
16	16	84.2	439	8	M85483	8.02e-03
17	16	84.2	433	19	T67045	8.02e-03
18	16	84.2	473	28	A1536458	8.02e-03
19	16	84.2	491	21	A1014144	8.02e-03
20	16	84.2	520	26	A148007	8.02e-03
21	16	84.2	520	27	A1452515	8.02e-03
22	16	84.2	541	28	A1498144	8.02e-03
23	16	84.2	543	18	AA800202	8.02e-03
24	16	84.2	703	41	AA031206	8.02e-03
25	16	84.2	802	41	A0344176	8.02e-03
26	15	78.9	198	36	AA090452	8.02e-03
27	15	78.9	274	31	H32634	1.26e-01
28	15	78.9	278	28	A1556258	1.26e-01
29	15	78.9	304	30	R63500	1.26e-01
30	15	78.9	304	38	A0045919	1.26e-01
31	15	78.9	313	21	A1030280	1.26e-01
32	15	78.9	338	22	A1080379	1.26e-01
33	15	78.9	355	28	A1548398	1.26e-01
34	15	78.9	358	38	B76117	1.26e-01
35	15	78.9	386	13	AA444411	1.26e-01
36	15	78.9	398	39	A0204173	1.26e-01
37	15	78.9	401	15	AA609146	1.26e-01
38	15	78.9	406	32	H88528	1.26e-01
39	15	78.9	411	32	N26651	1.26e-01
40	15	78.9	473	26	A1400570	1.26e-01
41	15	78.9	488	40	A0236754	1.26e-01
42	15	78.9	500	39	A0212680	1.26e-01
43	15	78.9	505	27	A1466753	1.26e-01
44	15	78.9	549	15	AA573121	1.26e-01
45	15	78.9	643	41	A0319314	1.26e-01

# ALIGNMENTS

RESULT	1	AA583407	551 bp	mrna	EST	08-SEP-1997
LOCUS		AA583407				
DEFINITION		nm42604.s1 NCI-CGAP_GCS Homo sapiens CDNA clone IMAGE:1086511 3'				
		similar to gb:L07395 PROTEIN PHOSPHATASE PPI-GAMMA CATALYTIC				
		SUBUNIT (HUMAN);, mRNA sequence.				
ACCESSION		AA583407				
NID		G2368016				
VERSION		AA583407.1				
KEYWORDS		EST.				
SOURCE		human.				
ORGANISM		Homo sapiens				
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
		Eutheria; Primates; Catarrhini; Homiidae; Homo.				
REFERENCE		1 (bases 1 to 551)				
AUTHORS		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
TITLE		National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
JOURNAL		Tumor Gene Index				
COMMENT		Unpublished (1997)				
		Contact: Robert Strausberg, Ph.D.				
		Tel: (301) 496-1550				
		Email: Robert.Strausberg@nih.gov				
		Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.				
		Emmert-Buck, M.D., Ph.D.				
		CDNA Library Preparation: Stratagene, Inc.				
		CDNA Library Arrayed by: Greg Lennon, Ph.D.				
		DNA Sequencing by: Washington University Genome Sequencing Center				
		Clone distribution: NCI-CGAP clone distribution information can be				
		found through the I.M.A.G.E. Consortium/LLNL at:				
		www.bio.llnl.gov/bdrp/image/image.html				
		Seq primer: -40m13 fwd. ET from Amerham				
		High quality sequence stop: 340.				
		Location/Qualifiers				
FEATURES						



```

source
1.551
/organism="Homo sapiens"
/Note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. Mixed
germ cell tumors. 5' adaptor sequence: 5' GAATTCGGCAGG
3' 3' adaptor sequence: 5' CCGGCGTTTCTTTTCTTTT 3'
Average insert size: 0.7 kb."
/db_xref="taxon:9606"
/clone_image="IMAGE:1086511"
/clone_lib="NCI_CGAP_GC5"
/tissue_type="germ cell tumor"
/lab_host="SOLR (kanamycin resistant)"

BASE COUNT      180 a      93 c      100 g      178 t
ORIGIN

Query Match      94.7% Score 18; DB 15; Length 551;
Best Local Similarity 100.0%; Pred. No. 2,43e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      521 GTTACAAATATTAAAGA 538
OY      1 GTTACAAATATTAAAGA 18

RESULT 2
LOCUS      AU037905      299 bp      mRNA      EST      29-MAR-1999
DEFINITION AU037905 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
discoideum cDNA clone SSB439, mRNA sequence.
ACCESSION      AU037905
NID      93984658
VERSION      AU037905.1 GI:3984658
KEYWORDS      EST.
SOURCE      Dictyostelium discoideum.
ORGANISM      Dictyostelium discoideum
Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE      1 (bases 1 to 299)
AUTHORS      Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mita,B.N., Pi,M., Saito,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development
JOURNAL      DNA Res. 5 (6), 335-340 (1998)
MEDLINE      99156227
COMMENT      On Jan 5, 1998 this sequence version replaced gi:2747224.

Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan
Email: d402huesakura.cc.tsukuba.ac.jp
PROJECT      "Dictyostelium discoideum cDNA project in Japan".
Location/Qualifiers
1. 299
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_image="SSB439"
/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slug"

BASE COUNT      129 a      39 c      16 g      109 t      6 others
ORIGIN

Query Match      89.5% Score 17; DB 25; Length 299;
Best Local Similarity 100.0%; Pred. No. 4.61e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      167 TTCTTAATATTGTAA 183
CP      18 TTCTTAATATTGTAA 2

RESULT 3

```

```

LOCUS      Z39615      253 bp      mRNA      EST      07-NOV-1994
DEFINITION HSCIEG052 normalized infant brain cDNA Homo sapiens cDNA clone
c-1eg05 3', mRNA sequence.
ACCESSION      Z39615
NID      9565410
VERSION      Z39615.1 GI:565410
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 253)
AUTHORS      Auitray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
Devignes,M.D., Duprat,S., Houigatte,R., Juneau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pletu,G., Pouliot,Y.,
Sebastiani-Kabackchis,C. and Tessier,A.
IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III. Sci. Vie 318 (2), 263-272 (1995)
95277534.

JOURNAL
MEDLINE
COMMENT

Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read. 19 T removed at sequence 5' end
Genexpress library idt: C; Genexpress_sequence_idt: alc-1eg05
Seq primer: (421)M13.universal.
Location/Qualifiers
1. 253
/organism="Homo sapiens"
/Note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI; sex=Female; dev_stage=3 months old;
Isolate=muscular atrophy patient; tissue_type=total
brain; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Souares, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S. in press."
/db_xref="taxon:9606"
/map="10 q11.2 or q22.1-q23"
/clone_image="c-1eg05"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"

BASE COUNT      99 a      33 c      37 g      82 t      2 others
ORIGIN

Query Match      84.2% Score 16; DB 8; Length 253;
Best Local Similarity 100.0%; Pred. No. 8.02e-03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      237 GTTACAAATATTAAAG 252
OY      1 GTTACAAATATTAAAG 16

RESULT 4
LOCUS      C20646      298 bp      mRNA      EST      23-OCT-1996
DEFINITION HUMGS0004648 Human adult (K.Okubo) Homo sapiens cDNA 3', mRNA
sequence.
ACCESSION      C20646
NID      91621756
VERSION      C20646.1 GI:1621756
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

```



REFERENCE 1 (bases 1 to 298)  
 AUTHORS Okubo, K.  
 TITLE BodyMap: human gene expression database  
 JOURNAL Unpublished (1995)  
 COMMENT On Apr 14, 1999 this sequence version replaced gi:716642.

Contact: Okubo, K.  
 Institute for Molecular and Cellular Biol  
 Osaka University  
 1-3, Yamada-oka, Suita, Osaka Pref. 565, Japan  
 Tel: 06-877-5111(ex 3315)  
 Email: kousaku@imcb.osaka-u.ac.jp  
 Human Gene Signature, 3'-directed cDNA sequence. We are not  
 submitting the same cDNA sequence redundantly to DBJ since 1993.  
 For the abundance information of clones with this sequence in this  
 library and as well as in other 3'-directed libraries, see  
 http://www.imcb.osaka-u.ac.jp/bodymap/. The sequences of the clones  
 represented by this GS sequences is also found there.

FEATURES  
 Location/Qualifiers

1..298  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="958B12: 5: 5p15.31-5p15.33"  
 /clone\_lib="Human adult (K.Okubo)"  
 /dev\_stage="adult"  
 93 a 48 c 45 g 102 t 10 others

Query Match 84.2%; Score 16; DB 36; Length 298;  
 Best Local Similarity 100.0%; Pred. No. 8.02e-03;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 134 CTTAATATTGTAC 149  
 16 CTTAATATTGTAC 1

RESULT 5  
 LOCUS A1383567 324 bp mRNA EST 18-MAR-1999  
 DEFINITION tc4se11.x1 Soares.total\_fetus\_Nb2HF8\_9w Homo sapiens cDNA clone  
 IMAGE:2067596 3', mRNA sequence.

ACCESSION A1383567  
 NID 94196348  
 VERSION A1383567.1 GI:4196348  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 324)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

REFERENCE 1 (bases 1 to 324)  
 AUTHORS NCI-CGAP  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 14, 1998 this sequence version replaced gi:1877535.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAG Consortium (info@image.lnl.gov) for further information.  
 Insert Length: 1047 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 313.

FEATURES  
 Location/Qualifiers

1..324  
 /organism="Homo sapiens"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from pooled 8-9 week  
 (total) fetus material with a Not I - oligo(dT) primer [5'  
 TCTTACCACTGAACTGGACGGCCGCTTATATTTTTTTTTTTT 3'].  
 Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 125 a 41 c 50 g 108 t  
 ORIGIN

Query Match 84.2%; Score 16; DB 26; Length 324;  
 Best Local Similarity 100.0%; Pred. No. 8.02e-03;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 240 GTTACCAATTTTAAAG 255  
 1 GTTACCAATTTTAAAG 16

RESULT 6  
 LOCUS A1474210 330 bp mRNA EST 30-MAR-1999  
 DEFINITION t983f08.x1 Soares.NbHMPu\_S1 Homo sapiens cDNA clone IMAGE:2115399  
 3', mRNA sequence.

ACCESSION A1474210  
 NID 94327255  
 VERSION A1474210.1 GI:4327255  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 330)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3188852.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Insert Length: 2163 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 322.

FEATURES  
 Location/Qualifiers

1..330  
 /organism="Homo sapiens"  
 /note="Organ: mixed (see below); Vector: pT73D-Pac  
 (Pharmacia) with a modified polylinker; Site 1: Not I;  
 Site 2: Eco RI; Equal amounts of plasmid DNA from three  
 normalized libraries (melanocyte 2bHM, pregnant uterus  
 NBHPV, and fetal heart NBH19W) were mixed, and ss circles  
 was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from pools of  
 5,000 clones made from the same 3 libraries. The pools  
 consisted of I.M.A.G.E. clones 260232-265223,  
 340488-345479, and 484488-489479."

/db\_xref="taxon:9606"  
 /clone="IMAGE:2067596"  
 /clone\_lib="Soares\_NbHMPu\_S1"  
 /tissue\_type="Pooled human melanocyte, fetal heart, and  
 pregnant uterus"  
 /lab\_host="DH10B"

BASE COUNT 129 a 44 c 52 g 105 t  
 ORIGIN

Query Match 84.2%; Score 16; DB 27; Length 330;  
 Best Local Similarity 100.0%; Pred. No. 8.02e-03;



Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 233 GTTACAAATATTAAAG 248  
 ||||||||||||||||  
 OY 1 GTTACAAATATTAAAG 16

RESULT 7  
 LOCUS T05047 346 bp mRNA EST 30-JUN-1993  
 DEFINITION EST02935 Fetal brain, Stratagene (cat#936206) Homo sapiens CDNA  
 clone HFB035, mRNA sequence.  
 ACCESSION T05047  
 NID 9316200  
 VERSION T05047.1 GI:316200  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 346)  
 AUTHORS Adams, M.D., Kerlavage, A.R., Fields, C. and Venter, J.C.  
 TITLE 3,400 expressed sequence tags identify diversity of transcripts  
 from human brain  
 JOURNAL Nature Genet. 4, 256-267 (1993)  
 MEDLINE 93364420  
 COMMENT

Contact: Adams, MD  
 The Institute for Genomic Research  
 932 Clopper Road, Gaithersburg, MD 20878  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: mdadams@tigr.org  
 Seq primer: M13-21.  
 Location/Qualifiers

FEATURES  
 SOURCE 1..346  
 /organism="Homo sapiens"  
 /note="Vector: LambdaZAP-II. 17-18 wk gestation, female;  
 oligo-dT + random primed CDNA synthesis; LambdaZAP-II  
 vector, 1.0kb average insert size."  
 /db\_xref="ATCC (Inhost):81784"  
 /db\_xref="taxon:9606"  
 /clone="HFB035"  
 /clone\_lib="Fetal brain, Stratagene (cat#936206)"  
 BASE COUNT 75 a 75 c 57 g 133 t 6 others  
 ORIGIN

Query Match 84.2%; Score 16; DB 8; Length 346;  
 Best Local Similarity 100.0%; Pred. No. 8.02e-03;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 188 TTCTTAATATTGTA 203  
 ||||||||||||||||  
 Cp 18 TTCTTAATATTGTA 3

RESULT 8  
 LOCUS AA988210 350 bp mRNA EST 27-JUL-1998  
 DEFINITION or98h07.s1 NCI\_CGAP\_Lus Homo sapiens CDNA clone IMAGE:1603933 3',  
 mRNA sequence.  
 ACCESSION AA988210  
 NID 93173574  
 VERSION AA988210.1 GI:3173574  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 350)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2150481.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
[www-bio.lnl.gov/bbrp/image/image.html](http://www-bio.lnl.gov/bbrp/image/image.html)

Insert Length: 1184 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 291.  
 Location/Qualifiers

FEATURES  
 SOURCE 1..350  
 /organism="Homo sapiens"  
 /note="Organ: lung; Vector: p773D-Pac (Pharmacia) with a  
 modified polylinker; 1st strand CDNA was prepared from  
 neuroendocrine lung carcinoma, and was then primed with a  
 Not I - oligo(dT) primer. Double-stranded CDNA was ligated  
 to Eco RI adaptors (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of the modified  
 p773 vector. Library is normalized. Library was  
 constructed by Bento Soares and M. Fatima Bonaldo."  
 /db\_xref="taxon:9606"  
 /map="21g"  
 /clone="IMAGE:1603933"  
 /clone\_lib="NCI\_CGAP\_Lus"  
 /tissue\_type="carcinoid"  
 /lab\_host="DH10B"  
 BASE COUNT 133 a 48 c 58 g 111 t  
 ORIGIN

Query Match 84.2%; Score 16; DB 21; Length 350;  
 Best Local Similarity 100.0%; Pred. No. 8.02e-03;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 238 GTTACAAATATTAAAG 253  
 ||||||||||||||||  
 OY 1 GTTACAAATATTAAAG 16

RESULT 9  
 LOCUS AA936264 369 bp mRNA EST 23-JUN-1998  
 DEFINITION or75c02.s1 Soares\_NFL\_T-GBC\_S1 Homo sapiens CDNA clone  
 IMAGE:1562498 3', mRNA sequence.  
 ACCESSION AA936264  
 NID 93094182  
 VERSION AA936264.1 GI:3094182  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 369)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 17, 1998 this sequence version replaced gi:2044387.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
 Insert Length: 843 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 337.  
 Location/Qualifiers

FEATURES







## ORIGIN

Query Match 84.2%; Score 16; DB 24; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 8.02e-03;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 253 GTTACAAATATTAAAG 268  
 |||||||  
 0y 1 GTTACAAATATTAAAG 16

## RESULT 12

LOCUS AI369619 410 bp mRNA EST 16-FEB-1999  
 DEFINITION qy70h02.x1 NCI\_CGAP\_Brn25 Homo sapiens cDNA clone IMAGE:2017395 3',  
 mRNA sequence.

ACCESSION AI369619  
 NID q1448372  
 VERSION AI369619.1 GI:4148372  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
 AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute / National Institute of Neurological  
 Disorders and Stroke, Brain Tumor Genome Anatomy Project  
 (CGAP/BTCAP), Tumor Gene Index  
 Unpublished (1998)

JOURNAL  
 COMMENT On Jan 14, 1998 this sequence version replaced gi:1797674.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
 Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLT at:  
 www.bio.lnlt.gov/dbtp/image/image.html

Insert Length: 838 Std Error: 0.00  
 Seq primer: -40UP from Gibco.

FEATURES  
 SOURCE Location/Qualifiers

1..410

/organism="Homo sapiens"

/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer (5',  
 TGTACCAATCTGAGTGGAGCGCCGATAGGTTTTTTTTTTTTTTT  
 T 3'); double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 Library is normalized, and was constructed by Bento  
 Soares and M. Fatima Bonaldo."

/db\_xref="taxon:9606"

/clone="IMAGE:2017395"

/clone\_lib="NCI\_CGAP\_Brn25"

/tissue\_type="anaplastic oligodendroglioma"

/lab\_host="DH10B"

BASE COUNT 150 a 64 c 67 g 128 t 1 others

## ORIGIN

Query Match 84.2%; Score 16; DB 26; Length 410;  
 Best Local Similarity 100.0%; Pred. No. 8.02e-03;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 240 GTTACAAATATTAAAG 255  
 |||||||  
 0y 1 GTTACAAATATTAAAG 16

## RESULT 13

LOCUS AA349916 414 bp mRNA EST 21-APR-1997  
 DEFINITION EST56978 Infant brain Homo sapiens cDNA 5' end similar to EST  
 containing LI repeat, mRNA sequence.

ACCESSION AA349916  
 NID 92002307  
 VERSION AA349916.1 GI:2002307  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
 AUTHORS Adams M.D., Kerlavage A.R., Fields C. and Venter J.C.  
 TITLE 3,400 expressed sequence tags identify diversity of transcripts  
 from human brain  
 Nature Genet. 4, 256-267 (1993)

JOURNAL  
 MEDLINE 93364420  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1394386.  
 Other ESTs: EST56977

Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3016699056  
 Fax: 3016699423

Email: arkerlavet@tigr.org  
 For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (http://www.tigr.org/tdb/hgi/hgi.html)  
 Seq primer: M13 Reverse.

FEATURES  
 SOURCE Location/Qualifiers

1..414

/organism="Homo sapiens"

/note="Organ: brain; Vector: lafm1d BA; Site\_1: HindIII;  
 Site\_2: NotI"

/db\_xref="ATCC (lnhost):150864"

/db\_xref="taxon:9606"

/clone\_lib="Infant brain"

/sex="Female"

/dev\_stage="Infant"

BASE COUNT 130 a 73 c 83 g 125 t 3 others

## ORIGIN

Query Match 84.2%; Score 16; DB 12; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 8.02e-03;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 103 TTACAAATATTAAAGA 118  
 |||||||  
 0y 2 TTACAAATATTAAAGA 17

## RESULT 14

LOCUS R07577 415 bp mRNA EST 05-APR-1995  
 DEFINITION y697b09.1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone  
 IMAGE:122657 3', mRNA sequence.

ACCESSION R07577  
 NID 9759500  
 VERSION R07577.1 GI:759500  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
 AUTHORS Haller L., Clark N., Dubuque T., Elliston K., Hawkins M.,  
 Holman M., Hultman M., Kucaba T., Le M., Lennon G., Matra M.,  
 Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,  
 Trevisakis E., Waterston R., Williamson A., Wohlmann P. and  
 Wilson R.



451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565

Email: mscaresblue.weeg.uiowa.edu  
The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult 12-day-Embryo library. cDNA library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics The following repetitive elements were found in this cDNA sequence: 1-32, >NR\_richlow\_complexity  
Seq primer: M13 Forward.  
Location/Qualifiers  
1. 424  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/note="Vector: pRT73D-Pac (Pharmacia) with a modified

1.415

```

libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture."
/db_xref=taxon:10116
/clone=UI-R-E0-da-f-02-0-UI"
/clone_1ib="UI-R-E0"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"

BASE COUNT      136 a      65 c      74 g      149 t
ORIGIN
Query Match          84.2%;   Score 16; DB 26; Length 424;
Best Local Similarity 100.0%; Pred. NO. 8.02e-03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Dn 139 GTTACAATATTAAAG 154
|||||TTTTTTTTTTT
|||TTTTTTTTTTT
yy 1 GTTACAATATTAAAG 16
```

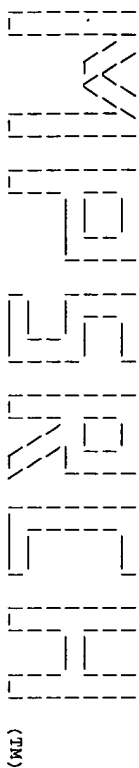
Search completed: Sat Nov 27 15:23:49 1999  
Job time : 225 secs.

Norway rat:  
*Rattus norvegicus*



**THIS PAGE BLANK (USPTO)**





(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*

Run on: Sat Nov 27 16:01:33 1999; Maspar time 110.23 Seconds

Tabular output not generated. 425.153 Million cell updates/sec

Title: >US-08-225-444-6  
Description: (1-20) from 5814619.seq  
Perfect Score: 20  
N.A. Sequence: 1 ACGTCCTGCAGTCGCGCG 20  
Comp: TGCAGCAGTCGCGCGCG

Scoring table: TABLE jmetric  
Gap 60

Mismatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: embl-est58  
1:em\_est10 2:em\_est11 3:em\_est17 4:em\_est18 5:em\_est2

Database: 6:em\_est9 7:em\_gss1  
genbank-est11

8:gb\_est1 9:gb\_est10 10:gb\_est11 11:gb\_est12 12:gb\_est13  
13:gb\_est14 14:gb\_est15 15:gb\_est16 16:gb\_est17  
17:gb\_est18 18:gb\_est19 19:gb\_est20 20:gb\_est21  
21:gb\_est22 22:gb\_est23 23:gb\_est24 24:gb\_est25  
25:gb\_est26 26:gb\_est27 27:gb\_est28 28:gb\_est29  
29:gb\_est30 30:gb\_est31 31:gb\_est32 32:gb\_est33 33:gb\_est34  
34:gb\_est35 35:gb\_est36 36:gb\_est37 37:gb\_gss1 38:gb\_gss2  
39:gb\_gss3 40:gb\_gss4 41:gb\_gss5 42:gb\_gss6

Statistics: Mean 5.681; Variance 1.084; scale 5.241

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description	Pred. No.
1	14	70.0	489 20	AA930978	vy79e12.r1 Stratagene	1.47e-02
2	14	70.0	493 16	AA663973	ac07c03.s1 Stratagene	1.47e-02
3	14	70.0	506 42	AA0399316	mgxb000812f1 CUG1 Rice	1.47e-02
4	14	70.0	516 18	AA786907	mg7b01a1.r1 Aspergillus	1.47e-02
5	14	70.0	603 25	AI295400	LP09018.5prime LP Dros	1.47e-02
6	13	65.0	147 35	AA051068	mg73d12.r1 Soares mous	3.46e-01
7	13	65.0	215 31	R97568	yg57g07.r1 Soares feta	3.46e-01
8	13	65.0	220 26	AI397401	fb22g03.y1 Zebratish W	3.46e-01
9	13	65.0	296 10	AA233357	zr29d07.r1 Stratagene	3.46e-01
10	13	65.0	303 20	AA882998	TENS0575 T. cruzi epim	3.46e-01

RESULT	1	AA930978	489 bp	MRNA	EST	23-APR-1998
LOCUS	1	AA930978	489 bp	MRNA	EST	23-APR-1998
DEFINITION	1	vy79e12.r1 Stratagene mouse macrophage (937306) Mus musculus cDNA				
ACCESSION	1	AA930978				
NID	1	93079411				
VERSION	1	AA930978.1	GI:3079411			
KEYWORDS	1	EST.				
SOURCE	1	house mouse.				
ORGANISM	1	Mus musculus				
REFERENCE	1	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	1	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
	1	(bases 1 to 489)				
	1	Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,				
	1	Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,				
	1	Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,				
	1	Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and				
	1	Waterston, R.				
TITLE	1	The Musku-HMT Mouse EST Project				
JOURNAL	1	Unpublished (1996)				
COMMENT	1	On Jan 19, 1998 this sequence version replaced gi:2285538.				

## ALIGNMENTS

Contact: Maria M/Mouse EST Project  
WashU-HMT Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1810  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:862758  
Seq primer: -28m13 rev1 ET from Amerham  
High quality sequence stop: 384.



FEATURES  
source

Location/Qualifiers  
1. 489  
/organism="Mus musculus"  
/note="Organ: blood; Vector: pBluescript SK-; Site: 1;  
EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. WEHI-3 cell line. Average insert size: 1.5 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3."  
/db\_xref="taxon:10090"  
/clone="IMAGE:1312462"  
/clone\_lib="Stratagene mouse macrophage (#937306)"  
/tissue\_type="macrophage"  
/dev\_stage="WEHI-3 cell line"  
/lab\_host="SOLR (kanamycin resistant)"

BASE COUNT  
ORIGIN

96 a 138 c 121 g 134 t

Query Match 70.0%; Score 14; DB 20; Length 489;  
Best Local Similarity 100.0%; Pred. No. 1,47e-02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 174 CTCGAGTCGCGTGC 187  
|||||  
6 CTCGAGTCGCGTGC 19

RESULT 2 AA663973 493 bp MRNA EST 12-NOV-1997

LOCUS ac07c03.s1 Stratagene lung (#937210) Homo sapiens cDNA clone  
DEFINITION IMAGE:855748 3' similar to gb:X57809 IG LAMBDA CHAIN C REGIONS  
(HUMAN); mRNA sequence.

ACCESSION AA663973  
NID 92617964  
VERSION AA663973.1 GI:2617964  
KEYWORDS EST  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 493)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,  
Kritman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Maita, M.,  
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU-MCI human EST Project  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1405126.

TITLE JOURNAL  
COMMENT

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewartson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40m3 fwd. ET from Amersham  
High quality sequence stop: 424.

FEATURES  
source

Location/Qualifiers  
1. 493  
/organism="Homo sapiens"  
/note="Organ: lung; Vector: pBluescript SK-; Site: 1;  
EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. normal lung. Average insert size: 1.0 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3."  
/db\_xref="taxon:9606"  
/clone="IMAGE:855748"  
/clone\_lib="Stratagene lung (#937210)"  
/sex="male"  
/dev\_stage="72 years"  
/lab\_host="SOLR cells (kanamycin resistant)"

BASE COUNT 84 a 121 c 164 g 124 t

ORIGIN

Query Match 70.0%; Score 14; DB 16; Length 493;  
Best Local Similarity 100.0%; Pred. No. 1,47e-02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 453 GCGACTGCAGACG 466  
|||||  
15 GCGACTGCAGACG 2

RESULT 3 AQ399316 506 bp DNA GSS 05-MAR-1999

LOCUS mgxb0008L21f CUGI Rice Blast BAC Library Magnaporthe grisea genomic  
DEFINITION clone.mgxb0008L21f, genomic survey sequence.

ACCESSION AQ399316  
NID 94370343  
VERSION AQ399316.1 GI:4370343  
KEYWORDS GSS.  
SOURCE Magnaporthe grisea.  
ORGANISM Magnaporthe grisea

REFERENCE 1 (bases 1 to 506)  
AUTHORS Yu, Y., Zhu, H., Boyd, C.A., Gaudette, B., Gayle, A., Kingsbury, R.,  
Phillips, K., Sasinowski, M., Wang, R.A. and Dean, R.A.  
A BAC End Sequencing Framework to Sequence the Magnaporthe grisea  
Genome  
Unpublished (1998)

TITLE JOURNAL  
COMMENT

CONTACT: Dean RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson University, SC 29634  
Tel: 864 656 5737  
Fax: 864 656 4293  
Email: rdeane@clemson.edu  
Seq primer: TAATACGACTCACTATAGG  
Class: BAC ends  
High quality sequence stop: 406.

FEATURES  
source

Location/Qualifiers  
1. 506  
/organism="Magnaporthe grisea"  
/strain="70-15"  
/note="Vector: pBACWICH; Site: 1: HindIII; Site: 2: HindIII;  
Rice blast is one of the most devastating fungal diseases  
of rice world wide. It is a filamentous ascomycete with  
a haploid genome (n=7) of approximately 40 Mbp. Rice  
blast is an important model fungal pathogen for studying  
numerous aspects of the fungal-host interaction. In  
order to facilitate genome wide analysis, a BAC library  
containing 9216 clones with an average insert size of 130  
kbp was constructed. This library represents greater  
than 25x genome coverage. High density colony filters  
are available upon request."  
/db\_xref="taxon:5138"  
/clone="mgxb0008L21f"  
/clone\_lib="CUGI Rice Blast BAC Library"  
/tissue\_type="Protoplasts"  
/lab\_host="E. coli DH10B"

BASE COUNT 143 a 131 c 96 g 136 t

Query Match 70.0%; Score 14; DB 42; Length 506;  
Best Local Similarity 100.0%; Pred. No. 1,47e-02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 116 CGACTGCAGACGCT 129  
|||||  
14 CGACTGCAGACGCT 1

RESULT 4



LOCUS AA786907 516 bp mRNA EST 31-JUL-1998  
 DEFINITION m7b01a1.r1 Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library Emericella nidulans cDNA clone m7b01a1 5', mRNA sequence.  
 ACCESSION AA786907  
 NID 92847237  
 VERSION AA786907.1 GI:2847237  
 KEYWORDS EST.  
 SOURCE Emericella nidulans.  
 ORGANISM Emericella nidulans.  
 Eukaryota; Fungi; Ascomycota; Euascomycetes; Plectomycetes; Eutriciales; Trichocomaceae; Emericella.  
 REFERENCE 1 (bases 1 to 516)  
 AUTHORS Kupfer, D., Gray, J., Hausner, J., Lai, H., Martin, W., Aramayo, R., Prade, R. and Roe, B.  
 TITLE An Aspergillus nidulans EST Database  
 JOURNAL Unpublished (1998)  
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2150589.  
 Other\_ESTs: m7b01a1.f1  
 Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu  
 Department of Chemistry and Biochemistry  
 Advanced Center for Genome Technology  
 University of Oklahoma  
 620 Parrington Oval, Norman, OK 73019, USA  
 Tel: 405 325 4912  
 Fax: 405 325 7762  
 Email: broe@ou.edu  
 We anticipate the future release of the cDNA clones to the Fungal Genetics Stock Center  
 .seq primer: SK  
 High quality sequence stop: 499.  
 FEATURES  
 source  
 1. 516  
 /organism="Emericella nidulans"  
 /strain="FGSC A26"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript 3' end of cDNA cloned into XhoI site of pBluescript"  
 /db\_xref="taxon:5072"  
 /map="21g"  
 /clone="m7b01a1"  
 /clone\_lib="Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library"  
 /tissue\_type="vegetative mycelia, asexual structures"  
 BASE COUNT 144 a 116 c 137 g 119 t  
 ORIGIN  
 Query Match 70.0%; Score 14; DB 18; Length 516;  
 Best Local Similarity 100.0%; Pred. No. 1.47e-02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 162 GCGAGCTGCAGGACG 175  
 Cp 15 GCGAGCTGCAGGACG 2  
 RESULT 5  
 LOCUS AI295400 603 bp mRNA EST 01-DEC-1998  
 DEFINITION LP09018.5prime LP Drosophila melanogaster larval-early pupal P0T2  
 ACCESSION AI295400  
 NID 93944807  
 VERSION AI295400.1 GI:3944807  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 603)  
 AUTHORS Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein, P., Lewis, S. and Rubin, G.M.  
 TITLE BDGP/HMI Drosophila EST Project  
 JOURNAL Unpublished (1997)

COMMENT On Jan 17, 1998 this sequence version replaced gi:2043824.  
 Contact: Harvey, D.  
 G. M. Rubin-Molecular and Cell Biology  
 University of California Berkeley  
 539 USA, Berkeley, CA 94720-3200, USA  
 Fax: 510 643 9947  
 Email: [http://www.fruitfly.org/EST\\_estefruitfly.berkeley.edu](http://www.fruitfly.org/EST_estefruitfly.berkeley.edu)  
 Sequence is the complete cDNA insert.  
 Plate: 90 row: B column: 6  
 High quality sequence stop: 475.  
 FEATURES  
 source  
 1. 603  
 /organism="Drosophila melanogaster"  
 /note="Organ: whole body; Vector: P0T2; Site\_1: EcoRI; Site\_2: XhoI; Sized fractionated cDNAs were directly ligated into P0T2. Plasmid cDNA library."  
 /db\_xref="taxon:7227"  
 /clone="LP09018"  
 /clone\_lib="LP Drosophila melanogaster larval-early pupal P0T2"  
 /sex="male and female"  
 /dev\_stage="larvae-pupae"  
 /lab\_host="DH5-alpha"  
 BASE COUNT 117 a 151 c 197 g 138 t  
 ORIGIN  
 Query Match 70.0%; Score 14; DB 25; Length 603;  
 Best Local Similarity 100.0%; Pred. No. 1.47e-02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 375 ACCTCCTGCAGTCG 388  
 Oy 1 ACCTCCTGCAGTCG 14  
 RESULT 6  
 LOCUS AA051068 147 bp mRNA EST 09-SEP-1996  
 DEFINITION mg73d12.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA  
 ACCESSION AA051068  
 NID 91530757  
 VERSION AA051068.1 GI:1530757  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 147)  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On May 5, 1995 this sequence version replaced gi:798442.  
 CONTACT: Maria M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: [mouseest@wustl.edu](mailto:mouseest@wustl.edu)  
 This clone is available royalty-free through LIND; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:263983  
 Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 135.  
 FEATURES  
 source  
 1. 147



Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314.286.1800  
Fax: 314.286.1810  
Email: estewartson.wustl.edu  
High quality sequence stops: 95  
Source: IMAGE Consortium, LNLN.  
This clone is available royalty-free through LNLN; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Seq primer: M13RP1  
High quality sequence stop: 95.  
Location/Qualifiers  
1. .215  
/organism="Homo sapiens"  
/note="Organ: Liver and Spleen; Vector: pT713D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer"

Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbraflsh@wustl.edu  
CDNA Library Preparation: Matthew Clark, CDNA Library Arrayed by:  
Matthew Clark, DNA Sequencing by: Washington University Genome  
Sequencing Center  
Clone distribution: Genome Systems, St. Louis,  
Missouri (web address: [www.genomesystems.com](http://www.genomesystems.com)) (email contact:  
[info@genomesystems.com](mailto:info@genomesystems.com)) and Research Genetics, Huntsville, Alabama  
(web address: [www.resgen.com](http://www.resgen.com)) (email contact: [info@resgen.com](mailto:info@resgen.com)) and  
Ressourcenzentrum Primatendatabank, Berlin, Germany (web address:  
[www.rzpd.de](http://www.rzpd.de))  
Trace considered overall poor quality  
Possible reversed clone: similarly on wrong strand  
Seq primer: T3 ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers  
1. 220  
/organism="Danio rerio"  
/note="Vector: pSPOR1, Site\_1: NotI, Site\_2: SalI, 1st











DB 333 TCCTGCAGTCGCG 345  
 |||||  
 4 TCCTGCAGTCGCG 16

RESULT 13  
 LOCUS AA887173 362 bp mRNA EST 30-MAR-1998  
 DEFINITION n251d04.s1 NCI\_CGAP\_Prl2 Homo sapiens cDNA clone IMAGE:1291303,  
 mRNA sequence.

ACCESSION AA887173  
 NID 93002281  
 VERSION AA887173.1 GI:3002281  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 362)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 17, 1998 this sequence version replaced gi:2044705.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D.,  
 Rodrigo F. Chuang, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: David B. Kitzman, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbp/image/image.html](http://www.bio.llnl.gov/dbp/image/image.html)

Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 361.

FEATURES  
 Location/Qualifiers  
 1..362  
 /organism="Homo sapiens"  
 /note="Vector: PAMP10; mRNA made from metastatic prostate  
 lesion of the bone, cDNA made by oligo-dt priming.  
 Non-directionally cloned. Size-selected on agarose gel,  
 average insert size 600 bp. Library made by D. Kitzman,  
 NIH."  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:1291303"  
 /clone\_1lb="NCI\_CGAP\_Prl2"  
 /sex="male"  
 /tissue\_type="metastatic prostate bone lesion"  
 /lab\_host="DH10B"  
 /lab\_host="DH10B"

BASE COUNT 79 a 78 c 113 g 92 t

Query Match 65.0%; Score 13; DB 20; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 3,46e-01;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 4 GCGACTGCAGGAC 16  
 |||||  
 Cp 15 GCGACTGCAGGAC 3

RESULT 14  
 LOCUS AA806951 366 bp mRNA EST 07-APR-1998  
 DEFINITION oc34a03.s1 NCI\_CGAP GCBI Homo sapiens cDNA clone IMAGE:1351564 3',  
 similar to gb:U15537\_rna2 PROTEIN-TYROSINE PHOSPHATASE 1C (HUMAN);,  
 mRNA sequence.

ACCESSION AA806951  
 NID 92876527  
 VERSION AA806951.1 GI:2876527

KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 366)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2150480.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
 Ph.D., Gerald Marti, M.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbp/image/image.html](http://www.bio.llnl.gov/dbp/image/image.html)

Insert Length: 1033 Std Error: 0.00  
 Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 135.

FEATURES  
 Location/Qualifiers  
 1..366  
 /organism="Homo sapiens"  
 /note="Vector: p773D-Pac (Pharmacia) with a modified  
 polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from human tonsillar cells enriched for  
 germinal center B cells by flow sorting (CD20+, IgD-),  
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
 primed with a Not I - oligo(dT) primer  
 [5'-NGTACCAATGCTGAAGGAGCGCGCCCACTTTTCTTTTCTTTT-  
 3']. Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified p773 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."  
 /db\_xref="taxon:9606"  
 /map="21q"  
 /clone\_image="IMAGE:1351564"  
 /clone\_1lb="NCI\_CGAP\_GCBI"  
 /tissue\_type="germinal center B cell"  
 /lab\_host="DH10B"  
 /lab\_host="DH10B"

BASE COUNT 64 a 99 c 103 g 100 t

Query Match 65.0%; Score 13; DB 18; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 3,46e-01;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 146 GCGACTGCAGGAC 158  
 |||||  
 Cp 15 GCGACTGCAGGAC 3

RESULT 15  
 LOCUS A1503863 372 bp mRNA EST 11-MAR-1999  
 DEFINITION vm26c07.x1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA  
 clone IMAGE:991308 3', mRNA sequence.

ACCESSION A1503863  
 NID 94401714  
 VERSION A1503863.1 GI:4401714  
 KEYWORDS EST.  
 SOURCE house mouse,  
 Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;



REFERENCE  
AUTHORS

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 372)  
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,  
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,  
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson,R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
On Mar 20, 1998 this sequence version replaced gi:2980012.

TITLE  
JOURNAL  
COMMENT

Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:563588

This clone was previously sequenced on the 5' end only, this new  
data is from the 3' end  
Seq primer: primer name ambiguous  
High quality sequence stop: 370.

FEATURES  
source

Location/Qualifiers  
1..372  
/organism="Mus musculus"  
/strain="B6D2 F1/J"  
/note="Organ: embryo; Vector: pSPORT; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally from mRNA prepared  
from 800 blastocysts. Primer: SalI(4T):  
5'-CGGTGACCGTCGACCGTTTCTTTTCTT-3'. CDNAs were  
cloned into the NotI/SalI sites of a pSPORT vector (Life  
Technologies). Two different size selections: B1 (larger  
inserts) and B3."  
/db\_xref="taxon:10090"  
/clone\_image:991308"  
/clone\_lib="Knowles Solter mouse blastocyst B1"  
/tissue\_type="blastocyst"  
/dev\_stage="embryo (pre-implantation)"  
/lab\_host="DH10B"  
/lab\_host="DH10B"

BASE COUNT 78 a 108 c 121 g 65 t  
ORIGIN

Query Match 65.0%; Score 13; DB 28; Length 372;  
Best Local Similarity 100.0%; Pred. No. 3,46e-01;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 340 TGCAGTCGCGTCG 352  
|||||  
QY 7 TGCAGTCGCGTCG 19

Search completed: Sat Nov 27 16:03:30 1999  
Job time : 117 secs.

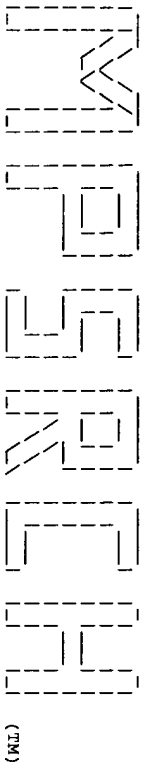






**THIS PAGE BLANK (USPTO)**





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSPrch\_nlp n.a. - n.a. Smith-Waterman search, using a protein database  
which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Nov 24 02:12:01 1999; MasPar time 153.50 Seconds  
1408.056 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-3  
Description: (1-660) from US09103287.seq  
Perfect Score: 3300  
N.A. Sequence: 1 ATTTAAAGTTTCGGATGACA.....GGCTTTATATGTTTATA 660  
Comp: TAAATTTCAAGCCTTCTGT.....CCCAATATATCAATATTT

Scoring table: TABLE bkttranslate2  
Gap 30

Mmatch STD : Dbase 0; Query 0

Searched: 179066 seqs, 163739223 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

sptrembl9  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mmc 8:sp-organelle  
9:sp-phase 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-invertebrate 14:sp-virus

Statistics: Mean 72.686; Variance 161.713; scale 0.449

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	313	9.5	454	2	067373	UDP-N-ACETYLURAMATE-A	4,64e-23
2	269	8.2	803	2	084767	UDP-N-ACETYLURAMATE-A	1.56e-16
3	167	5.1	373	8	035905	VARIP.	4.44e-03
4	168	5.1	872	5	P90523	PUTATIVE TRANSCRIPTION	3.41e-03
5	165	5.0	101	5	025801	COMPLETE GENE MAP OF P	7.52e-03
6	164	5.0	161	14	088439	COMPLETE GENOME.	9.76e-03
7	163	5.0	4350	5	077336	MAL3P3.18 PROTEIN.	1.64e-02
8	162	4.9	960	5	025802	FRAMESHIFT.	4.57e-02
9	158	4.8	149	8	047514	NADH DEHYDROGENASE SUB	2.74e-02
10	160	4.8	277	14	098460	GENOME, PARTIAL SEQUEN	2.74e-02
11	160	4.8	2708	5	015791	CHLOROQUINE RESISTANCE	2.74e-02
12	160	4.8	2742	5	015801	CHLOROQUINE RESISTANCE	2.74e-02
13	160	4.8	2819	5	015792	STRAIN HB3 CG2 (CG2).	1.25e-01
14	154	4.7	1532	2	077332	MAL3P3.16 PROTEIN.	1.25e-01
15	152	4.6	88	2	066201	SIMILAR TO GROES PROTE	2.05e-01
16	152	4.6	239	5	025815	COMPLETE GENE MAP OF P	2.05e-01

C	17	153	4.6	310	2	050846	IMMUNOGENIC PROTEIN P3	1.60e-01
C	18	151	4.6	398	2	050911	CONSERVED HYPOTHETICAL	2.62e-01
C	19	151	4.6	535	10	004472	SIMILAR TO ZEA MAVS PE	2.62e-01
C	20	153	4.6	2343	6	062730	FACTOR VIII.	1.60e-01
C	21	147	4.5	153	14	065150	LECTIN-LIKE PROTEIN.	6.90e-01
C	22	150	4.5	337	3	013953	HYPOTHETICAL 39.1 KD P	3.34e-01
C	23	147	4.5	347	2	050905	CONSERVED HYPOTHETICAL	6.90e-01
C	24	148	4.5	350	8	033572	KINETOPLAST APOCYTOCHR	5.43e-01
C	25	149	4.5	367	2	051731	THYMIDINE KINASE (TDK)	4.26e-01
C	26	149	4.5	391	5	025768	ASPARAGINE-RICH ANTIGE	4.26e-01
C	27	147	4.5	534	5	043989	HOMEBBOX-CONTAINING PR	6.90e-01
C	28	147	4.5	726	2	P72664	HYPOTHETICAL 86.5 KD P	6.90e-01
C	29	147	4.5	1225	5	015784	HISTIDINE KINASE C.	6.90e-01
C	30	144	4.4	91	5	025616	COMPLETE GENE MAP OF P	1.41e+00
C	31	144	4.4	404	8	095946	Y61BBB.1 PROTEIN.	1.41e+00
C	32	144	4.4	402	8	095946	MITOCHONDRION TRANSFER	1.41e+00
C	33	145	4.4	470	8	002696	HYPOTHETICAL 54.7 KD P	1.11e+00
C	34	145	4.4	791	5	000841	PUTATIVE TRANSCRIPTION	1.11e+00
C	35	145	4.4	812	5	018717	C49H3.5 PROTEIN.	1.11e+00
C	36	145	4.4	1306	5	077273	EG:66A1.2 PROTEIN.	1.11e+00
C	37	146	4.4	1655	5	024754	MASTERMIND.	8.77e-01
C	38	144	4.4	1690	5	044929	MICROTUBULE BINDING PR	1.41e+00
C	39	146	4.4	1799	5	018220	COSMID C26E6.	8.77e-01
C	40	146	4.4	1802	5	018219	COSMID C26E6.	8.77e-01
C	41	143	4.3	130	14	036833	MA-P17 (FRAGMENT).	1.78e+00
C	42	143	4.3	993	3	048673	NI5B.	1.78e+00
C	43	143	4.3	1982	5	021440	WVO-3 PROTEIN.	1.78e+00
C	44	143	4.3	2343	6	018806	FACTOR VIII.	1.78e+00
C	45	143	4.3	2657	5	077380	MAL3P6.15 PROTEIN.	1.78e+00

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	454 AA.
AC	067373	067373;		
AD	01-AUG-1998	(TREMBLERL. 07, CREATED)		
DT	01-AUG-1998	(TREMBLERL. 07, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBLERL. 08, LAST ANNOTATION UPDATE)		
DE	UDP-N-ACETYLURAMATE-ALANINE LIGASE.			
GN	MURC.			
OS	AQUIFEX AEOLICUS.			
OC	BACTERIA; AQUIFICALES; AQUIFICACEAE; AQUIFEX.			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-VF5.			
RA	DECKERT G., WARREN P.V., GAESTERLAND T., YOUNG W.G., LENOX A.L.,			
RA	GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,			
RA	FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;			
RT	"The complete genome of the hyperthermophilic bacterium Aquifex			
RT	aeolicus".			
RL	NATURE 392:353-358(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-VF5.			
RA	DECKERT G., WARREN P.V., GAESTERLAND T., YOUNG W.G., LENOX A.L.,			
RA	GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,			
RA	FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;			
DR	SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.			
KW	EMBL: AE000736; G2983764; -			
LI	LIGASE.			
SO	SEQUENCE 454 AA; 50893 MW; 120799CE CRC32;			
Query Match	9.5%; Score 313; DB 2; Length 454;			
Best Local Similarity	39.4%; Pred. No. 4,64e-23;			
Matches	117; Conservative 54; Mismatches 121; Indels 5; Gaps 5;			
DB	H N V Y N A L A A T G V A L E L G V S F			
DC	831 HCAVYAVGNTAVAGCYTNGCNCNG-GNGTNCNTNGARYTNGGNGTWSMT 889			
OY	121 CCAATACAGTTTAAAGCATGACCTGATTAATTCGATTAATTTGGAACAGCTAGATGT 180			
Ot	H T V L N A L A V I A L V I * R S * M L			



[illegible][illegible]



Dd N N I N N I N N I N N M N M N  
Dc 8395 AATTAATAAATAAAATTAATAAATAAATAAATAAGAAAATAATGAT 846  
|||:::|::|::|::|::|::|::|::|::|:  
Cp 595 AATTACCTCCATTAATTAACAACGCCTTTCATTTGTCTAATACATTAAT 536



[illegible]

	Query Match	4.8%	Score 158;	Length 149;
	Best Local Similarity	31.1%;	Pred. No. 4; DB-07;	
Matches	36;	Conservative	34;	Mismatches 50; Indels 0; Gaps 0
Dd	F V V L I F F S G I F S L L T Y F C S M			
Dc	152 AATYGTGCTGTTNATHTTYYTWSNGGNATHHWTYSVNTYNACNTAATYTGGWMSNA 211			
Cp	342 ACTTCATTATGCGATATTTCTTGCGTGCCTGTCACATGTAGACACTATTTCTTGGA 283			
Ct	F F M W I F L S C C V N C S T N F S W M			
Dd	S N F I F Y Y N Y F F F F S Y L F L V S F			
Dc	212 TGSMNAATGTTATHTTYATAAYTAATYATYTTTYYTTTYSWYNTTNTYTNCTNSMT 271			
Cp	282 TGCTGTGATATCATCTCAACAATACCTGATTCGAAATGTAGTTGATTTGAAACGACGT 223			
Ct	V C I I I Y N N L I C N C S F I E T T F			
Dd	272 TY 273			
Dc	I :			
Cp	222 TT 221			
Ct	:			
RESULT	10			
ID	098460	PRELIMINARY;	PRT:	277 AA.
AC	098460:			
Df	01-FEB-1997 (TREMBLE REL. 02, CREATED)			
Dt	01-FEB-1997 (TREMBLE REL. 02, LAST SEQUENCE UPDATE)			
Dt	01-NOV-1998 (TREMBLE REL. 08, LAST ANNOTATION UPDATE)			
De	GENOME, PARTIAL SEQUENCE.			
Gn	A408L.			
Oc	PARAMECIUM BURSARIA CHLORELLA VIRUS 1 (PBCV-1).			
Os	VIRUSES; DSDNA. VIRUSES, NO RNA STAGE; PHYCODNAVIRIDAE; PHYCODNAVIRUS.			
Rn	[1]			
Rp	SEQUENCE FROM N.A.			
Rx	MEDLINE; 96400190.			
Ry	KUTISH G.F., LI Y., LU Z., FURUTA M., ROCK D.L., VAN ETYEN J.L.;			
Rt	"Analysis of 76 kb of the chlorella virus PBCV-1 330-kb genome: map			
Rt	positions 182 to 258."			
Rl	VIROLOGY 223:303-317(1996).			
Rl	EMB.; 042580; G1620080; -.			
So	SEQUENCE 277 AA; 30772 MW; F51A62A1 CRC32;			
	Query Match	4.8%; Score 160; DB 14; Length 277;		
	Best Local Similarity	33.6%;	Pred. No. 2; 74e-02;	
Matches	47;	Conservative	33;	Mismatches 60; Indels 0; Gaps 0
Dd	N T F L I C Y F S L C S D T S R I N L S			
Dc	11 GAAAVACMTVTYNAHTGYATYTTWMNTNGTWSNGAVACNMNMGNATHAAYTNW 70			
Cp	348 GCACAACTCTTTATGTGATATTTCTTGCTGCTGTCGTCATGTAGACACTATTTCT 289			
Ct	N I F F M W I F L S C C V N C S T N F S			
Dd	N I V F I I T T D T I M I S D S F N S M			
Dc	71 SNAATATGTTTATTHATTHACNCNGVACNATHATGATMGATWSNNTYAAYSNA 130			
Cp	288 CTGTGATGCTGTATATCATCTACACATATCTGATTTGCAATGTAGTTGATTGCAA 229			
Ct	W M V C I I I Y N N L I C N C S F I E T			
Dd	V L D F I K			
Dc	131 TGTNYTNGAATYTAATHAR 150			
Cp	228 CGAGCTTAAACACCACCAA 209			
Ct	T F N T T K			
RESULT	11			
ID	015791	PRELIMINARY;	PRT:	2708 AA.
AC	015791:			
Df	01-JAN-1998 (TREMBLE REL. 05, CREATED)			
Dt	01-JAN-1998 (TREMBLE REL. 05, LAST SEQUENCE UPDATE)			
Dt	01-NOV-1998 (TREMBLE REL. 08, LAST ANNOTATION UPDATE)			

	Query Match	4.8%	Score 158;	Length 149;
	Best Local Similarity	31.1%;	Pred. No. 4; DB-07;	
	Matches	36; Conservative	34; Mismatches	50; Indels 0; Gaps 0
Dd	F V V L I F F S G I F S L L T Y F C S M			
Dc	152 AATYGTGCTGTTNATHTTYYWNGGNANHTTWYSNYVTNACNTAATYTYGYWSNA 211			
Cp	342 ACTTCATTATGGCATATTTCTTGCGTGCCTGTCATATGAGCACTAATTTCTTGGA 283			
Ct	F F M W I F L S C C V N C S T N F S W M			
Dd	S N F I F Y Y N Y F F F F S Y L F L V S F			
Dc	212 TGSMNAATGTTATHTTAYTAATAVATYTTTYYTTWTYSWYNTTNTYTNCTNSMT 271			
Cp	282 TGCTGTGATATACATCNCACATACTGATTCGAAATGATAGTTGATTAAGAAGCGCT 223			
Ct	V C I I I Y N N L I C N C S F I E T T F			
Dd	272 TY 273			
Dc	I :			
Cp	222 TT 221			
Ct	:			
RESULT	10	PRELIMINARY;	PRT:	277 AA.
ID	098460			
AC	098460:			
Df	01-FEB-1997 (TREMBLE REL. 02, CREATED)			
Dt	01-FEB-1997 (TREMBLE REL. 02, LAST SEQUENCE UPDATE)			
Dt	01-NOV-1998 (TREMBLE REL. 08, LAST ANNOTATION UPDATE)			
De	GENOME, PARTIAL SEQUENCE.			
Gn	A408L.			
Oc	PARAMECIUM BURSARIA CHLORELLA VIRUS 1 (PBCV-1).			
Os	VIRUSES; DSDMA VIRUSES, NO RNA STAGE; PHYCODNAVIRIDAE; PHYCODNAVIRUS.			
Rn	[1]			
Rp	SEQUENCE FROM N.A.			
Rx	MEDLINE; 96400190.			
Ra	KUTISH G.F., LI Y., LU Z., FURUTA M., ROCK D.L., VAN ETZEN J.L.;			
Rt	"Analysis of 76 kb of the chlorella virus PBCV-1 330-kb genome: map			
Rt	positions 182 to 258."			
Rl	VIROLOGY 223:303-317(1996).			
Rl	EMB.; 042580; G1620080; -.			
So	SEQUENCE 277 AA; 30772 MW; F51A62A1 CRC32;			
	Query Match	4.8%; Score 160; DB 14; Length 277;		
	Best Local Similarity	33.6%;	Pred. No. 2; 74e-02;	
	Matches	47; Conservative	33; Mismatches	60; Indels 0; Gaps 0
Dd	N T F L I C Y F S L C S D T S R I N L S			
Dc	11 GAAAVACMTVTYNATHGTATYATTTWMSNTTGWMSGAVACNMNMGNATHAAYTNW 70			
Cp	348 GCACAACCTCTTATATGATATTTCTTGCTGCTGTCATATGAGCACTAATTTCT 289			
Ct	N I F F M W I F L S C C V N C S T N F S			
Dd	N I V F I I T T D T I M I S D S F N S M			
Dc	71 SNAATATGTTTATTHATTHACNCGNAGYACNATHATGATMGATWSNNTYAAYSNA 130			
Cp	288 CTGATGCTGATATATCATATCTACATATGATTCGATTTGCAATGATGATTTCAATGAA 229			
Ct	W M V C I I I Y N N L I C N C S F I E T			
Dd	V L D F I K			
Dc	131 TGTNYTNGATYTAATHAR 150			
Cp	228 CGAGCTTAAACACCACCAA 209			
Ct	T F N T T K			
RESULT	11	PRELIMINARY;	PRT:	2708 AA.
ID	015791			
AC	015791:			
Df	01-JAN-1998 (TREMBLE REL. 05, CREATED)			
Dt	01-JAN-1998 (TREMBLE REL. 05, LAST SEQUENCE UPDATE)			
Dt	01-NOV-1998 (TREMBLE REL. 08, LAST ANNOTATION UPDATE)			

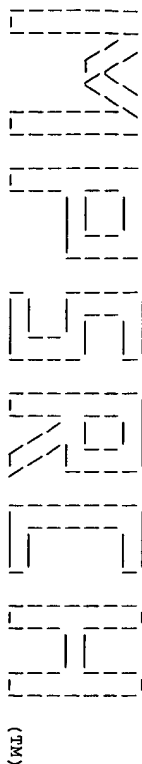


RESULT	14				
ID	077332	PRELIMINARY;	PRT;	1532 AA.	
AC	077332:				
DT	01-NOV-1998 (TREMBLEL. 08,				
DT	01-NOV-1998 (TREMBLEL. 08,				
DT	01-NOV-1998 (TREMBLEL. 08,				
DE	MALP3.16 PROTEIN.				
GN	MALP3.16.				
OS	PLASMODIUM FALCIPARUM.				
OC	EUFARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MUNGALL K., LAWSON D., BARRELL B.;				
RL	SUBMITTED (SEP-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.				
DR	EMBL: Z98547; E1332592;				
SO	SEQUENCE 1532 AA; 166754 MW; B3F0FC43 CRC32;				
Query Match		4.7%;	Score 154;	DB 5;	Length 1532;
Best Local Similarity		33.5%;	Pred. No. 1,256-01;		
Matches	58;	Conservative 43;	Mismatches 72;	Indels	0









(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Merch\_nu n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 12:12:54 1999; MasPar time 1159.65 Seconds  
1333.584 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-3  
Description: (1-660) from US09103287.seq  
Perfect Score: 660  
N.A. Sequence: 1 AATTAAAGATTCGATGACA.....GCGTTTAAATGTTATATA 660  
Comp: TAAATTTCAGCGCTACTGT.....CGCAAAATATACAAATAT

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2683791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: embl-est58  
1:em\_est10 2:em\_est11 3:em\_est17 4:em\_est18 5:em\_est2

Database: genbank-est111  
8:gb\_est1 9:gb\_est10 10:gb\_est11 11:gb\_est12 12:gb\_est13

13:gb\_est14 14:gb\_est15 15:gb\_est16 16:gb\_est17  
17:gb\_est18 18:gb\_est19 19:gb\_est20 20:gb\_est21  
21:gb\_est22 22:gb\_est23 23:gb\_est24 24:gb\_est25  
25:gb\_est26 26:gb\_est27 27:gb\_est28 28:gb\_est29  
29:gb\_est30 30:gb\_est31 31:gb\_est32 32:gb\_est33  
33:gb\_est34 34:gb\_est35 35:gb\_est36 36:gb\_est37  
37:gb\_est38 38:gb\_est39 39:gb\_gss3 40:gb\_gss4  
41:gb\_gss5 42:gb\_gss6

Statistics: Mean 10.899; Variance 2.639; scale 4.130

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	67	10.2	252	17	AA754459	97SN1787 Rice Immature 5.01e-61
2	52	7.9	252	17	AA754459	97SN1787 Rice Immature 6.06e-39
3	50	7.6	247	17	AA754458	97SN1784 Rice Immature 4.21e-36
4	36	5.5	2275	20	AF034173	AF034173 Human mRNA (T 3.18e-16
5	35	5.3	2275	20	AF034173	AF034173 Human mRNA (T 5.79e-16
6	35	4.2	439	38	B50349	CIT-HSP-351C15.TV CIT- 1.26e-07
7	28	3.8	727	26	AU005149	AU005149 Bombyx mori P 2.23e-04
8	25	3.8	768	28	A1525758	PT1.3_05_D12.r tumor1 2.23e-04
9	25	3.6	208	31	D56970	HUM24OG07B Clontech hu 2.36e-03
10	24					

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
11	24	3.6	273	21	C94261	C94261 Dictyostellium d 2.36e-03
12	24	3.6	276	38	A0060967	CIT-HSP-2355J16.TR CIT 2.36e-03
13	24	3.6	282	20	C91268	C91268 Dictyostellium d 2.36e-03
14	24	3.6	294	21	C93878	C93878 Dictyostellium d 2.36e-03
15	24	3.6	339	38	B98968	CIT-HSP-2280A9.TR CIT- 2.36e-03
16	24	3.6	360	23	A1111775	UT-R-Y0-me-9-06-0-01.s 2.36e-03
17	24	3.6	361	21	C90922	C90922 Dictyostellium d 2.36e-03
18	24	3.6	388	38	A0064726	HS-2214_B1.C08.MF CIT 2.36e-03
19	24	3.6	395	38	A0059895	CIT-HSP-2348N2.TR CIT- 2.36e-03
20	24	3.6	428	39	A0108899	CIT-HSP-2378A17.TR CIT 2.36e-03
21	24	3.6	452	14	C28215	C28215 Rice callus CDN 2.36e-03
22	24	3.6	479	24	A1220783	G902607.x1 Soares-plac 2.36e-03
23	24	3.6	513	25	A0040013	A0040013 Dictyostellium 2.36e-03
24	24	3.6	545	41	A0345845	RC111-12611.TJ RPI11 2.36e-03
25	24	3.6	562	38	A0045846	RC111-13515.TJ RPI11 2.36e-03
26	24	3.6	580	42	A0385581	RC111-13315.TV RPI11 2.36e-03
27	23	3.5	198	20	AA912400	O199606.s1 NCI.CGAP_PN 2.33e-02
28	23	3.5	274	21	C92129	C92129 Dictyostellium d 2.33e-02
29	23	3.5	277	10	AA229467	nc45f02.r1 NCI.CGAP_Pr 2.33e-02
30	23	3.5	349	19	T93588	Y17603.s1 StrataGene 2.33e-02
31	23	3.5	366	19	R22316	YH2611.s1 Soares-plac 2.33e-02
32	23	3.5	369	26	AU003619	AU003619 Bombyx mori P 2.33e-02
33	23	3.5	401	20	AA923684	O187H11.s1 NCI.CGAP_K1 2.33e-02
34	23	3.5	404	31	R97582	Y959406.r1 Soares-feta 2.33e-02
35	23	3.5	408	24	A1216565	GM42A11.x1 NCI.CGAP_Lu 2.33e-02
36	23	3.5	413	34	W95464	ze02c08.s1 Soares-feta 2.33e-02
37	23	3.5	417	21	A1027445	0065409.x1 Soares-plac 2.33e-02
38	23	3.5	422	20	R43706	Y919612.s1 Soares-inf 2.33e-02
39	23	3.5	428	14	AA462029	Z998903.s1 NCI.CGAP_GC 2.33e-02
40	23	3.5	436	25	A1343670	QP06510.x1 NCI.CGAP_K1 2.33e-02
41	23	3.5	463	16	AA688091	nv14G04.s1 NCI.CGAP_Pr 2.33e-02
42	23	3.5	469	33	N62159	Y62607.s1 Soares-mit 2.33e-02
43	23	3.5	617	39	A0110139	CIT-HSP-2376C5.TR CIT- 2.33e-02
44	23	3.5	620	34	W51835	zc49e10.s1 Soares-sene 2.33e-02
45	23	3.5	692	37	B78401	TJ2H21TF TAYU Arabidop 2.33e-02

## ALIGNMENTS

RESULT 1  
LOCUS AA754459 252 bp mRNA  
DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII CDNA Library Oryza sativa  
ACCESSION AA754459  
NID 92801165  
VERSION AA754459.1 GI:2801165  
KEYWORDS EST.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.  
1 (bases 1 to 252)  
Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.  
Large-scale Sequencing Analysis of ESTs from Rice Immature Seed Unpublished (1998)  
On Jan 14, 1998 this sequence version replaced gi:1197457.

TITLE JOURNAL  
COMMENT  
Contact: Eun M.Y.  
Department of Cytogenetics  
National Inst. of Agri. Sci. and Tech, RDA  
Suwon, Kyunggi-do, Korea  
Tel: 82 331 290 0301  
Fax: 82 331 290 0307  
Email: myeun@n20.osti.re.kr  
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr  
Seq primer: M13 Reverse Primer.  
Location/Qualifiers  
1. 252  
/organism="Oryza sativa"



```

/cultivar="Milyang23"
/node="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPIII
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"

```

Query Match	Score	DB	Length
Best Local Similarity	14.48	Pred. No. 6.06e-39,	
Matches	31	Conservative 108;	Mismatches 72; Indels 5; Gaps 5;
Db	37	HGNM5VHNCBTRGTHDCCKNMVSMTWTGTYMMBVNSDMHMYEBVNTKYDVGNHTRCS	96
Oy	437	TGTGAATTTTGGATGCAATTAGAGAAA-TACTG-GCCGATTAACGATACAGATTA-	493
Db	97	RMRVETMAHYHYTNCBVENNNDYMHMBBYMBVBTGCTCTMCMCBHYNTKCTASGHT	156
Oy	494	ATTGATAAATGGAAGTCATCGTTAATTAATGAAGATTCTATTA-ATGATTTAGAACA	552
Db	157	STNDVKSSTNTGVTBSYDKSMHGYWCSBBAVYHTKYSTTRATRSYTCVRRYCYMMWTK	216
Oy	553	ATTGATTAAGCTG-TTGTTTTATTATATGGGCGACAGTGATATTCAAAAATTACCAAAATG	611
Db	217	KVYKKYHVBGCBHBDSCKTMWNTKHKVMTSTD	252
Oy	612	CATATTTAGATAATTAGCATGAAAAATGCGTTT	647
RESULT	3	AA754458	247 bp mRNA EST 20-JAN-1998
LOCUS			

ACCESSION	AA754458
NID	92801164
VERSION	AA754458.1
KEYWORDS	GI:2801164
SOURCE	EST
ORGANISM	Oryza sativa. Oryza sativa. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE	1 (Bases 1 to 247)
AUTHORS	Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.
TITLE	Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL	Unpublished (1998)
COMMENT	On Jan 14, 1998 this sequence version replaced gi:1797455.

Contact: Eun M.Y.  
Department of Cytogenetics  
National Inst. of Agril. Sci. and Tech, RDA

Tel: 82 331 290 0301  
Fax: 82 331 290 0307  
Email: myeun@sun20.asti.re.kr  
Submitted by Baek Hye Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr  
Seq primer: M13 Reverse Primer.  
Location/Qualifiers  
1. 247  
/organism="Oryza sativa"  
/cultivar="MilYang23"  
/note="Vector: pBluescript SK(+), Site\_1: EcoRI, Site\_2: XhoI, Directional cDNA library inserted into lambda ZapR vector at 5' end with EcoRI and 3' end with Xho I site."



```

/db_xref="taxon:4530"
/map="6"
/clone="97SN1784"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT      7 a      16 c      21 g      34 t      169 others
ORIGIN
Query Match      7.6%; Score 50; DB 17; Length 247;
Best Local Similarity 12.9%; Pred. No. 4,21e-36;
Matches 29; Conservative 105; Mismatches 89; Indels 2; Gaps 2;

3 DCTMNTVYRGCCCBAMNKRHTHMTBMCVVRVGTNNKNGKRTTMDSCDNAAC 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28 TCACATATTCACAAATTCAGCATTAAGACTGCTGTTGATGTGATGTGAGTGTG 87
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 RYTBMYIARAKYKYGIBYBSMNVDNTGTGTGKTTVAVHSGMNRGNSVYVWETAY 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
88 TTATGATCAGCTCTGCTGTCACAAATATGTTGACATA-CAGTTTAAATGATTAAGTG 146
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 CDYHYEDRANHHVDITCTNDRCYNTASDNTSKTKVETGDKTDSGGGCKRKTY 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 147 TAAATGCGATTAGTATTATAGACAGCTAGATGTTCAATATTAAGAAGCATTAGAAA 206
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 GSSBYBRGCVNVVYTTSMWTDKSTKM-BSMDNSRSRVHYGRMM 226
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 207 CGTTGGTGGTGAACGTCGTTTCATGAACTACATGCA 251
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT      4
LOCUS      9A754458      247 bp      mRNA      EST      20-JAN-1998
DEFINITION      97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
ACCESSION      AA754458
VERSION      G2801164
KEYWORDS      Oryza sativa.
SOURCE      Oryza sativa.
ORGANISM      Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
REFERENCE      1 (bases 1 to 247)
AUTHORS      Nahm,B.H., Kim,W.Y., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
      Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
      Lee,M.C. and Eun,M.Y.
      Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
      Unpublished (1998)
      On Jan 14, 1998 this sequence version replaced gi:1797455.

TITLE      Contact: Eun M.Y.
JOURNAL      Department of Cyto genetics
COMMENT      National Inst. of Agri. Sci. and Tech, RDA
      Suwon, Kyunggido, Korea
      Tel: 82 331 290 0301
      Fax: 82 331 290 0307
      Email: myeun@suno.asti.re.kr
      Submitted by Beek Hie Nahm, Dept of Biological Science, Myongji
      University, Yongin, Korea. 449-728 bhnam@bioserver.myongji.ac.kr
      Seq primer: M13 Reverse Primer.
      Location/Qualifiers
          1. 247
              /organism="Oryza sativa"
              /cultivar="Milyang23"
              /note="vector: pBluescript SK(+); Site 1: EcoRI; Site 2:
              XhoI; Directional cDNA library inserted into lambda ZAPII
              vector at 5' end with EcoRI and 3' end with Xho I site."
              /db_xref="taxon:4530"
              /map="6"
              /clone="97SN1784"
              /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
FEATURES
      source

```

			/tissue_type="Immature Seed"
			/dev_stage="5 days after pollination"
			/lab_host="E. coli SOLR"
BASE COUNT	7 a	16 c	21 g
ORIGIN			34 t 169 others
Query Match	7.6%	Score 50;	DB 17; Length 247;
Best Local Similarity	14.7%;	Pred. No. 4.2le-36;	
Matches	33;	Conservative 106;	Mismatches 82; Indels 4; Gaps 4.
Db	22 KHTHMTBBCCVRRVGTTTNGKHNSGRTTWNCSSNAHCRIVBMYRYRSKYGYCTBY	81	
Cp	460 TCTAATTCATCAAAAATTTCACATAAGATACACGACTC-GCTTTACTTAACCTT-TCT	403	
Db	82 YSMWYDNTGGTGKTVNHSGMNNRCNSVYYVBIAVCYDBHYBDRAHHVDPRCT	141	
Cp	402 GCAAAATCATTTAAAAANG-CTGTGTCTCGAGAAGAAGTGCTGGAATACGCAC	344	
Db	142 NDRGICANT-A-SDNSTSKTYVTGDKTDSGCCGCKRKRYTGSBSYBRGVNVMTTS	200	
Cp	343 AACTCTTTAATGGATATTTCTTCGTCGCTCATATGTACACATAATTTCTTGG	284	
Db	201 MWTDSFKBMSDMRSRRSVYGHMMBKRGMSRNMDTKWT	245	
Cp	283 ATGTGTGCATATCATCTACATAACTTAGATTGCCAATTGTACT	239	
RESULT	5		
LOCUS	AF034173	2275 bp	mRNA EST
DEFINITION	AF034173 Human mRNA (Tripodis and Ragousis)		Homo sapiens CDNA
ACCESSION	AF034173		
NID	g2707735		
VERSION	AF034173.1	GI:2707735	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Euhayota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homindae; Homo.		
TITLE	1 (bases 1 to 2275) Tripodis, N. and Ragousis, J. Generation of a transcription map in the region immediately centromeric to human MHC across the 6p21-2-6p21.3 Chromosomal boundary Unpublished (1997)		
JOURNAL COMMENT	On Jan 19, 1998 this sequence version replaced gi:2045115.		
FEATURES	Contact: Tripodis, Nikos Division of Medical and Molecular Genetics Guys Hospital 7th floor, Guy's Tower, London SE1 9RT, UK Email: nikos@nki.nl. Location/Qualifiers 1..2275 /organism="Homo sapiens" /db_xref="taxon:9606" /map="6p21.3" /clone="ntcon2 contig" /clone_1bp="Human mRNA (Tripodis and Ragousis)" BASE COUNT 438 a 619 c 470 g 599 t 149 others		
ORIGIN			
Query Match	5.5%;	Score 36;	DB 20; Length 2275;
Best Local Similarity	15.3%;	Pred. No. 3.18e-17;	
Matches	19;	Conservative 66;	Mismatches 37; Indels 2; Gaps 2.
Db	1555 MKKCWKRRKKYKYSTYKYSRMYYTTYTYWCOCCTSUKASCAMMRMGYGRSS	1584	
Cp	350 CTGCACAACAATCTT-TTATGCGATATTTCTTCGTGCTG-TGCAATGTAGACACTAAT	283	
Db	1585 RSYWNYGSMGCGCMKRKRYRYSWTGWTKTWTWYMMSMKRWMTYTITWTYRTIKTW	1644	
Cp	292 TTTCCTTGGAATGATATCATCTACAACTGATGATTCGCAATGATGATTCATT	233	







ORIGIN

Query Match 3.8%; Score 25; DB 26; Length 727;  
Best Local Similarity 83.8%; Pred. No. 2.23e-04;  
Matches 31; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 468 GTTAATTAATAATTTACTAATGATATTAACAA 504  
517 GTTAATTAATAAGATTCATTAATGATATTAACAA 553

RESULT 9 A1525758 768 bp mRNA EST 18-MAR-1999

LOCUS PT1.3.05.D12.r tumor1 Homo sapiens cDNA 5', mRNA sequence.

ACCESSION A1525758

VERSION 94439893

KEYWORDS A1525758.1 GI:4439893

SOURCE EST.

ORGANISM human.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS 1 (bases 1 to 768)

Huang, G.M., Ng, W., Farkas, J., Chen, L., Liang, H.A., Gordon, D., Jun  
Yu, J. and Hood, L.

Prostate Cancer Expression Profiling by cDNA Sequencing Analysis

Unpublished (1999)

On May 18, 1998 this sequence version replaced gi:3137680.

COMMENT

Contact: Guyang Matthew Huang

Leroy Hood

University of Washington

Department of Molecular Biotechnology, Box 357730, University of

Washington, Seattle, WA 98195

Tel: 5106280100

Fax: 5106280108

Email: huangm@yahoo.com.

FEATURES Location/Qualifiers

1..768

/organism="Homo sapiens"

/note="Organ: Prostate; Vector: pBluescript; Directional

cDNA library was constructed using lambda ZP II kit

(Stratagene). mRNA was extracted from a frozen prostate

tumor tissue (Mayo Clinics)."

/db\_xref="taxon:9606"

/map="6p21.3"

/clone\_lib="tumor1"

BASE COUNT 205 a 163 c 191 g 160 t 49 others

ORIGIN

Query Match 3.8%; Score 25; DB 26; Length 768;

Best Local Similarity 71.7%; Pred. No. 2.23e-04;

Matches 38; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 526 CTTTATCAAAAACCAATTTTGGGCAACAATGCCCTCNAAGCTGAAT 578

53 CTTTATCCGTAATTTGAATTTTGAGCATTAATGTCATCCGATCTTAAT 1

RESULT 10 D56970 208 bp mRNA EST 28-AUG-1995

LOCUS HUM240G07B Clontech human aorta polyA+ mRNA (#6572) Homo sapiens

DEFINITION CDNA clone GEN-240G07 5', mRNA sequence.

ACCESSION D56970

VERSION 9963592

KEYWORDS D56970.1 GI:963592

SOURCE EST.

ORGANISM human.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 208)

Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M.,

Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H.,  
Takaichi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y.,  
Maekawa, H., Shin, S. and Nakamura, Y.  
Fujiwara et al. (1995)  
Unpublished (1995)  
On May 5, 1995 this sequence version replaced gi:797884.

Contact: Tsutomu Fujiwara  
Otsuka Gen Research Institute  
Otsuka Pharmaceutical Co., Ltd  
463-10 Kagasuno Kawachino-cho, Tokushima, Tokushima, 771-01 Japan  
Tel: 0886-65-2888  
Fax: 0886-37-1035  
Insert Length: 542 Std Error: 0.00  
High quality sequence stop: 271.

FEATURES

source

1..208

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Clontech human aorta polyA+ mRNA (#6572)"

BASE COUNT 61 a 30 c 26 g 75 t 16 others

ORIGIN

Query Match 3.6%; Score 24; DB 31; Length 208;

Best Local Similarity 78.9%; Pred. No. 2.36e-03;

Matches 30; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 108 TWTGATACGTAATAATTTGGATTAATTTGGAAT 145

429 TATCTTATGTAATAATTTTGGATCAATAGAGAAAT 466

RESULT 11 C94261 273 bp mRNA EST 15-JUN-1998

LOCUS C94261 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium

DEFINITION dictyodeum cDNA clone SSK635, mRNA sequence.

ACCESSION C94261

VERSION 93218876

KEYWORDS C94261.1 GI:3218876

SOURCE EST.

ORGANISM Dictyostelium discoideum.

REFERENCE Eukaryota; Dictyostelidia; Dictyostelium.

1 (bases 1 to 273)

Yoshino, R., Morio, T. and Tanaka, Y.

Unpublished (1997)

On Jan 14, 1998 this sequence version replaced gi:1797461.

COMMENT

Contact: Hideko Urushihara

Institute of Biological Sciences

University of Tsukuba

3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan

Email: d402huesakura.cc.tsukuba.ac.jp

PROJECT = Dictyostelium discoideum cDNA project in Japan.

FEATURES Location/Qualifiers

1..273

/organism="Dictyostelium discoideum"

/strain="AX4"

/db\_xref="taxon:44689"

/map="6"

/clone="SSK635"

/clone\_lib="Dictyostelium discoideum SS (H.Urushihara)"

/dev\_stage="slug"

BASE COUNT 100 a 41 c 27 g 105 t

ORIGIN

Query Match 3.6%; Score 24; DB 21; Length 273;

Best Local Similarity 73.1%; Pred. No. 2.36e-03;

Matches 38; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Db 213 AATATTAATGTTTTTAAATAAATTAATTAATCATTTTAAAT 264







```

DEFINITION   CIT-HSP-2280A9.TR CIT-HSP Homo sapiens genomic clone 2280A9,
              genomic survey sequence.
ACCESSION    B98968
NID          93026778
VERSION      B98968.1 GI:3026778
KEYWORDS     GSS.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Homidae; Homo.
REFERENCE    1 (bases 1 to 339)
AUTHORS      Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
              Golden,K., Berry,K., Granger,D., Sun,E., Wible,C., Shizuya,H.,
              Simon,M. and Venter,J.C.
              Use of a random BAC End Sequence Database for Sequence-Ready Map
              Building
              Building
              Unpublished (1997)
JOURNAL
COMMENT      Contact: Mark Adams
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: mdadams@tigr.org
              Clones are available from Research Genetics (info@resgen.com). BAC
              end search page:
              http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
              Seq primer: M13 Reverse
              Class: BAC ends.
FEATURES
  source      Location/Qualifiers
              1..339
                /organism="Homo sapiens"
                /note="Vector: pBelobAC11; site_1: HindIII; site_2:
                HindIII"
                /db_xref="taxon:9606"
                /clone="2280A9"
                /clone_1id="CIT-HSP"
                /sex="Male"
                /cell_type="Sperm"
BASE COUNT   123 a      61 c      38 g      117 t
ORIGIN
Query Match      3.6%; Score 24; DB 38; Length 339;
Best Local Similarity 71.4%; Pred. No. 2,36e-03;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
Db 61 TTTATTAAACCACTATCAACACCATCAATATCTTTAATGCTCATCAATTT 116
    ||| || ||||| || ||||| ||||| ||||| ||||| ||||| |||||
Cp 239 TTTCATTGAACGACGTTTAACACCAACCAAGCTTCTAATGCTCTTATATATT 184
    ||| || ||||| || ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: Sat Nov 27 12:32:26 1999  
 Job time : 1172 secs.



**THIS PAGE BLANK (USPTO)**



\*\*\*\*\*  
WISREH  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Nov 24 00:49:16 1999; Maspar time 10.17 Seconds

Tabular output not generated. 449.783 Million cell updates/sec

Title: >US-09-103-287-4

Description: (1-215) from US09103287.pep

Perfect Score: 1495

Sequence: 1 FKSDSDIYAGIFQITDKGTA.....GDIQKLNAYLDMKGNMF 215

Scoring table: PAM 150

Gap 11

Searched: 170751 seqs, 2126608 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 31.787; Variance 165.505; scale 0.192

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description	Pred. No.
1	1434	95.9	215	39	M89199		Partial sequence of t	8.46e-105	
2	1434	95.9	437	39	M87771		UDP-N-acetylmuramate:	8.46e-105	
3	651	43.5	422	33	M55120		Streptococcus pneumoniae	3.14e-41	
4	255	17.7	46	34	W77686		Staphylococcus aureus	5.35e-11	
5	235	15.7	455	22	W20606		H. pylori cytoplasmic	3.65e-08	
6	237	15.2	286	22	W24585		H. pylori cytoplasmic	3.65e-08	
7	227	15.2	286	22	W20102		H. pylori cytoplasmic	3.65e-08	
8	137	9.2	335	33	M55117		Streptococcus pneumoniae	9.70e-02	
9	137	9.2	450	26	W29454		Streptococcus pneumoniae	9.70e-02	
10	137	9.2	450	26	W68551		S. pneumoniae MurD pr	1.33e-01	
11	135	9.0	1786	23	W20436		P. falciparum liver s	1.44e-01	
12	104	7.0	187	22	W20436		H. pylori protein.	1.44e-01	
13	104	7.0	283	29	M55463		H. pylori ORF 05ae302	1.44e-01	
14	104	7.0	293	29	M55463		H. pylori ORF 05ae202	1.44e-01	
15	104	7.0	432	22	W20733		H. pylori cell envelope	1.44e-01	
16	95	6.4	418	8	R49247		PcK.	5.26e+01	

17	95	6.4	418	4	R22025	A. chrysogenum phosph	5.26e+01
18	94	6.3	417	12	R67000	Staphylococcus epider	6.06e+01
19	94	6.3	565	35	W71554	Helicobacter polypept	6.06e+01
20	94	6.3	677	38	M85608	Secreted protein clon	6.06e+01
21	92	6.2	411	21	W01558	TYPI protein.	8.03e+01
22	91	6.1	218	3	R20033	Rat brain glutathione	9.24e+01
23	90	6.0	39	W89183	S. pneumoniae GlnM OR	1.06e+02	
24	90	6.0	417	4	R22095	Phosphoglycerate kina	1.06e+02
25	89	6.0	417	31	W54358	Phosphoglycerate kina	1.22e+02
26	90	6.0	459	39	W89182	S. pneumoniae GlnM po	1.06e+02
27	90	6.0	573	2	P70345	Portion of B fragment	1.06e+02
28	89	6.0	953	30	W56011	Recombinant botulinum	1.22e+02
29	88	5.5	218	1	R24927	Glutathione S-transfe	1.40e+02
30	88	5.5	241	2	R24185	Bovine RSV strain A 5	1.40e+02
31	88	5.9	456	13	R74996	E. maxima Em70-1 anti	1.40e+02
32	88	5.9	478	17	R95956	Eukaryotic cell growt	1.40e+02
33	88	5.9	15281	9	R44929	T. niivum Cyclosporin	1.40e+02
34	87	5.8	367	32	W56617	Serpulina hydrosenter	1.85e+02
35	86	5.8	370	35	W71543	Helicobacter polypept	1.85e+02
36	87	5.8	497	16	R81548	Hemolysin.	1.85e+02
37	86	5.8	524	30	W52579	Rabbit excitatory ami	1.85e+02
38	86	5.8	528	22	W20908	H. pylori inner membr	1.85e+02
39	86	5.8	613	15	R74632	QETR ethylene respons	1.85e+02
40	87	5.8	665	27	W27660	Streptococcus pneumon	1.61e+02
41	86	5.8	694	1	R04107	DNA-binding protein G	1.85e+02
42	87	5.8	3144	11	R58777	Protein encoded by Hu	1.61e+02
43	87	5.8	3144	22	W09871	Human huntingtin.	1.61e+02
44	87	5.8	3144	26	W36887	Previously undescribe	1.61e+02
45	87	5.8	3144	29	W44742	Human huntingtin prot	1.61e+02

## ALIGNMENTS

RESULT 1  
ID W89199 standard; Protein: 215 AA.  
AC W89199;  
DE 17-MAR-1999 (first entry)  
DE Partial sequence of the MurC polypeptide.  
KW MurC gene; UDP-N-acetylmuramate:L-alanine ligase; MurC polypeptide;  
KW bacterial; infection; H. pylori; cancer; gastric; gastritis; vaccine;  
KW Immunogen; drug; genetic immunisation.  
OS Staphylococcus aureus.  
PN EP-889123-A2.  
PE 07-JUN-1999.  
PF 26-JUN-1998; 305064.  
PR 03-JUL-1997; US-052720.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PI Burnham MKR, Wallis NG;  
PI WPI: 99-062655/06.  
DR N-PEDB; V80065.  
PT New isolated MurC polypeptide from Staphylococcus aureus and related  
PT nucleic acid - useful in diagnosis, treatment and prevention of  
PT bacterial infections  
PS Claim 1; Page 5; 39pp; English.  
CC The invention relates to a UDP-N-acetylmuramate:L-alanine ligase  
CC (MurC polypeptide) encoded by the S. aureus MurC gene. Host cells  
CC containing an expression system comprising the MurC gene can be used for  
CC the recombinant production of the polypeptide. Agonists or the MurC  
CC polypeptide are used to treat conditions requiring increased activity or  
CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or  
CC competitive polypeptide are useful for inhibiting the polypeptide e.g.  
CC bacterial (especially S. aureus) infections. They are also useful against  
CC Helicobacter pylori infections and related cancers, ulcers and gastritis.  
CC The antibacterial agents are useful to treat in-dwelling devices for  
CC infection prevention or generally as wound treatments to prevent adhesion  
CC of bacteria to matrix proteins. The MurC polypeptide is also useful for  
CC diagnosing or prognosing a (susceptibility to) disease, for raising  
CC antibodies, to identify modulators or specific receptors, in rational  
CC drug design and as an immunogen for vaccines. The MurC gene sequences are  
CC useful in antisense/ribozyme therapeutics; to detect mutant MurC gene;  
CC for chromosomal mapping; to determine bacterial serotype; and for genetic  
CC immunisation. The present sequence represents a partial sequence of the



CC Mucr polypeptide.  
SQ Sequence 215 AA;

95.9%: Score 1434; DB 39; Length 215;  
Best Local Similarity 96.3%; Pred. No. 8 466-105;

Query Match  
Matches 207; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 1 fkdssdiiyagndiqitckgtatfvydgefydhflspqygdhtvlnalavialstyelekldv 60  
1 FKDSDDIYAIQIQTIDKGTAVDVYDGEFYDHFLSPOQGDHTVNLNLAVALIATYLEKLDV 60  
Qy 1 fkdssdiiyagndiqitckgtatfvydgefydhflspqygdhtvlnalavialstyelekldv 60  
Db 61 tnikaletfgvkrirrfnetliangviyddahhpresatidtarckypkhevavfqp 120  
61 TNIKALETFGCVKRRFNETTIANOVIVDDYAHHPRESATIDTARKKYPKHEVAVFQP 120  
Qy 61 tnikaletfgvkrirrfnetliangviyddahhpresatidtarckypkhevavfqp 120  
Db 121 htfsrtqaflnfeaslskadrvficeifgsirensagalltqdlidkigasslinedin 180  
121 HTFSRTQAFLENFASLSKADRVFICEIFGSIRENTGALTITODLIDKIGASLINEDSIN 180  
Qy 121 HTFSRTQAFLENFASLSKADRVFICEIFGSIRENTGALTITODLIDKIGASLINEDSIN 180  
Db 181 vlegfdnnavlvmgagdiqkignayldkigmknaf 215  
181 VLEGF DNNAVLFMGAGDIQKIGNAYLDKIGMKNAF 215  
Qy 181 vlegfdnnavlvmgagdiqkignayldkigmknaf 215

RESULT 2  
ID W87771 standard; Protein: 437 AA.  
AC W87771.

DT 17-MAR-1999 (first entry)

DE UDP-N-acetylmuramate:L-alanine ligase (Mucr polypeptide).

KW Mucr gene; UDP-N-acetylmuramate:L-alanine ligase; Mucr polypeptide;

bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;

immunogen; drug; genetic immunisation.

OS Staphylococcus aureus.

PN EP-889123-A2.

PD 07-JAN-1999.

PF 26-JUN-1998; 305064.

PR 03-JUL-1997; US-052720.

PA (SMIK ) SMITHKLINE BEECHAM CORP.

PI (SMIK ) SMITHKLINE BEECHAM PLC.

DR Burnham MKR, Wallis NG;

DR WPI: 99-062655/06.

DR N-PSDB: V99650.

PT New isolated Mucr polypeptide from Staphylococcus aureus and related

PT nucleic acid - useful in diagnosis, treatment and prevention of

PT bacterial infections

PS Claim 1: Page 4: 39pp; English.

CC The present sequence represents a UDP-N-acetylmuramate:L-alanine ligase

CC (Mucr polypeptide) encoded by the S. aureus Mucr gene. Host cells

CC containing an expression system comprising the Mucr gene can be used for

CC the recombinant production of the polypeptide. Agonists or the Mucr

CC polypeptide are used to treat conditions requiring increased activity or

CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or

CC competitive polypeptide are useful for inhibiting the polypeptide e.g.

CC bacterial (especially S. aureus) infections. They are also useful against

CC Helicobacter pylori infections and related cancers, ulcers and gastritis.

CC The antibacterial agents are useful to treat in-dwelling devices for

CC infection prevention or generally as wound treatments to prevent adhesion

CC of bacteria to matrix proteins. The Mucr polypeptide is also useful for

CC diagnosing or prognosing a (susceptibility to) disease, for raising

CC antibodies: to identify modulators or specific receptors; in rational

CC drug design and as an immunogen for vaccines. The Mucr gene sequences are

CC useful in antisense/Ribozyme therapeutics; to detect mutant Mucr gene;

CC for chromosomal mapping; to determine bacterial serotype; and for genetic

CC immunisation.

SQ Sequence 437 AA;

Query Match 95.9%: Score 1434; DB 39; Length 437;

Best Local Similarity 96.3%; Pred. No. 8 466-105;  
Matches 207; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 223 fkdssdiiyagndiqitckgtatfvydgefydhflspqygdhtvlnalavialstyelekldv 282  
1 FKDSDDIYAIQIQTIDKGTAVDVYDGEFYDHFLSPOQGDHTVNLNLAVALIATYLEKLDV 60

Db 283 tnikaletfgvkrirrfnetliangviyddahhpresatidtarckypkhevavfqp 342  
61 TNIKALETFGCVKRRFNETTIANOVIVDDYAHHPRESATIDTARKKYPKHEVAVFQP 120  
Qy 61 tnikaletfgvkrirrfnetliangviyddahhpresatidtarckypkhevavfqp 342  
Db 343 htfsrtqaflnfeaslskadrvficeifgsirensagalltqdlidkigasslinedin 402  
121 HTFSRTQAFLENFASLSKADRVFICEIFGSIRENTGALTITODLIDKIGASLINEDSIN 180  
Qy 121 HTFSRTQAFLENFASLSKADRVFICEIFGSIRENTGALTITODLIDKIGASLINEDSIN 180  
Db 403 vlegfdnnavlvmgagdiqkignayldkigmknaf 437  
181 VLEGF DNNAVLFMGAGDIQKIGNAYLDKIGMKNAF 215  
Qy 181 vlegfdnnavlvmgagdiqkignayldkigmknaf 437

RESULT 3  
ID W51120 standard; Protein: 422 AA.  
AC W51120.

DT 02-OCT-1998 (first entry)

DE Streptococcus pneumoniae SP0070 protein.

KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;

KM detection; pneumonia; otitis media; meningitis.

OS Streptococcus pneumoniae.

PN W09818930-A2.

PD 07-MAY-1998.

PF 30-OCT-1997; U19422.

PR 31-OCT-1996; US-029960.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Choi GH, Hromocky J A, Johnson LS, Kunsch CA;

DR WPI: 98-272224/24.

DR N-PSDB: V27381.

PT Nucleic acid encoding antigenic peptide(s) from Streptococcus

PT pneumoniae - or their epitope-containing fragments, useful in

PT protective or therapeutic vaccines, and for diagnosis

PS Claim 11: Page 73; 11pp; English.

CC The present sequence represents a protein from Streptococcus pneumoniae.

CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein

CC can be useful in vaccines for inducing protective antibodies against

CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.

CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid

CC are used to detect Streptococcus infection (by usual hybridisation or

CC amplification methods), also for isolating Streptococcus genes or their

CC allelic variants. The protein can be used similarly for detecting specific

CC antibodies in standard immunoassays, especially for diagnosing or

CC monitoring infections. Antibodies which bind the protein are used to

CC detect corresponding antigens, to purify the protein and for passive

CC immunisation (optionally coupled to a toxin). Vaccines are administered,

CC e.g. by injection, orally or through the skin, typically at 0.01-1000

CC (especially 10-300) mu g/ml per dose.

SQ Sequence 422 AA;

Query Match 43.5%: Score 651; DB 33; Length 422;

Best Local Similarity 44.0%; Pred. No. 3 14e-41;

Matches 92; Conservative 54; Mismatches 60; Indels 3; Gaps 3;

Db 206 egndfvasdlirsstgftvfhfgnlgfhtpfrhmatavaglllytagfainl 265  
3 DSDDIYAIQIQTIDKGTAVDVYDGEFYDHFLSPOQGDHTVNLNLAVALIATYLEKLDV 62

Qy 3 DSDDIYAIQIQTIDKGTAVDVYDGEFYDHFLSPOQGDHTVNLNLAVALIATYLEKLDV 62

Db 266 vrehlktfagvkrirrfnetliangviyddahhpresatidtarckypkhevavfqp 325  
63 IKKALETFGCVKRRFNETTIANOVIVDDYAHHPRESATIDTARKKYPKHEVAVFQP 122

Qy 63 IKKALETFGCVKRRFNETTIANOVIVDDYAHHPRESATIDTARKKYPKHEVAVFQP 122

Db 326 ftrtialldfahbnadavylagiygsarevghdvkvedlanlknkhyqvtvsnvs 365  
123 FSRITQAFLENFASLSKADRVFICEIFGSIRENTGALTITODLIDKIGASLINEDSIN 180

Qy 123 FSRITQAFLENFASLSKADRVFICEIFGSIRENTGALTITODLIDKIGASLINEDSIN 180

Db 386 plidhnavyvmgagdiqyeysf-erl 413  
181 VLEGF DNNAVLFMGAGDIQKIGNAYLDKIGMKNAF 209

RESULT 4  
ID W77686 standard; Protein: 46 AA.



AC W27686; 1998 (first entry)  
 DT 30-OCT-1998  
 DE Staphylococcus aureus protein of unknown function.  
 KW Staphylococcus aureus protein; immune response induction; eye infection;  
 KW antibody production; T-cell immune response; gastrointestinal infection;  
 KW respiratory infection; inhibitor; bacterial infection; cardiac infection;  
 KW central nervous system; kidney infection; urinary tract infection;  
 KW antimicrobial compound identification; broad spectrum antibiotic;  
 KW therapy.  
 OS Staphylococcus aureus.  
 PN EP-841394-A2.  
 PD 13-MAY-1998.  
 PF 24-SEP-1997; 307485.  
 PR 24-SEP-1996; US-027032.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Burnham MKR, Hodgson JE, Knowles DJC,  
 PI Loretto MA, Nicholas RO, Pratt JM, Relchard RW, Rosenberg M,  
 PI Ward JN; 96-252940/23.  
 DR WPI: 96-252940/23.  
 DR New nucleic acid sequences from Staphylococcus aureus WCHU29 -  
 DR NT-PSDB: V53479.  
 PT useful in vaccines and for treatment of bacterial infections of e.g.  
 PT respiratory tract and central nervous system  
 PS Claim 11; Page 329; 390pp; English.  
 CC This sequence represents a Staphylococcus aureus protein of unknown  
 CC function, and is encoded by a DNA sequence of the invention.  
 CC The DNA sequences were isolated from Staphylococcus aureus WCHU29  
 CC (NCMB 40711). Host cells containing the DNA sequences are used to  
 CC produce polypeptides or fragments. The proteins are used in the treatment  
 CC of disease, for inducing an immune response by administering them, to  
 CC produce antibody and/or T-cell immune response. Antagonists of the  
 CC proteins are used for the inhibition of bacterial polypeptides.  
 CC Conditions which may be treated include bacterial infections, especially  
 CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,  
 CC urinary tract, skin, bones and joints. The proteins can also be used to  
 CC identify antimicrobial compounds which are broad spectrum antibiotics,  
 CC especially useful in the treatment of H. pylori infection.  
 SQ Sequence 46 AA;

[illegible]

involved in outer membrane or cell wall biosynthesis. The protein may be used in a vaccine to prevent or treat *H. pylori* infection or to identify *H. pylori* polypeptide binding compounds, useful as potential *H. pylori* life cycle activators or inhibitors. The genomic sequence of *H. pylori* (ATCC 55679) was determined from overlapping contigs generated by mechanically shearing the bacterial DNA. The sequences were analysed for ORF of at least 180 nucleotides, and the predicted coding regions defined by computer evaluation. To identify likely *H. pylori* antigens for vaccine development, the amino acid sequences predicted from various ORF were analysed for significant homology to other known or exported membrane proteins. Having identified and determined the sequences of interest, particular regions can be isolated from *H. pylori* by PCR amplification for recombinant polypeptide production, e.g. in *E. coli* hosts.

	Query Match	15.7%	Score 235;	DB 22	Length 455;
	Best Local Similarity	28.6%	Pred. No. 9, 34e-09;		
	Matches 44;	Conservative	42;	Mismatches 60;	Indels 8; Gaps 7
Db	241	ekkdlyngqylkld-gepytsf-eiknlgafywgigehmatna-slaalsaldelnee	297		
Qy	3	DSDDLYIAQIFQITDKGTAVDYVVDGEFFDHFISPOYGHTYVLMALAVAIISYLEKLDVTN	62		
Db	298	lrmnlfnkqikrfdilqkndlllddyahphteigstikksariyanlntqekilwiw	357		
Qy	63	IKALETFGCVYKRRENETTIANQVAVDDYVAHHPRIISATIDTARK-KYPRK-E-VVAVF	118		
Db	358	gahkysrlmdnleekfcflehcdtlllpyysa	391		
Qy	119	OPHTSRKQAFLEPAES-LSKADKVFISCTFGS	151		

RESULT	6
ID	W24585 standard; Protein; 286 AA.
AC	W24585;
DT	04-AUG-1997 (first entry)
DE	H. pylori cytoplasmic protein, 11253.aa.
KM	Transmembrane; cytoplasmic; cell envelope; flagella; transport;
KW	secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
KM	activator; inhibitor; bacterial life cycle; vaccine; immune;
KW	detection; antisense; inhibition.
OS	Helicobacter pylori.
FH	Key Location/Qualifiers
FT	misc_difference 2 /note= "encoded by AAS"
FT	misc_difference 133 /note= "encoded by TYA"
FT	misc_difference 166 /note= "encoded by RCC"
FN	WO9719098-A1.
PD	29-MAY-1997.
PF	15-NOV-1996; U18542.
PR	17-NOV-1995; US-561469.
PA	(ASTR ) ASTRA AB.
PI	Smith DH:
DR	WPII.97-298052/27.
N-PDB:	T77043.
PT	Helicobacter pylori nucleic acid sequences and related proteins -
PT	used for diagnostics and therapeutics
PS	Claim 18: Page 131: 235pp; English.
CC	This sequence represents an H. pylori cytoplasmic protein involved in
CC	outer membrane or cell wall biosynthesis. This sequence showed
CC	homology to N-acetylmuramate-Alanine ligase.
CC	Helicobacter pylori has been strongly linked to chronic gastritis and
CC	duodenal ulcer disease. The nucleic acid sequences of the invention
CC	are used to evaluate compounds, especially activators or inhibitors of
CC	bacterial life cycle, for the ability to bind an H. pylori nucleic acid
CC	sequence. The nucleic acid sequences, and corresponding proteins, are
CC	also useful for generating vaccines for immunising subjects against H.
CC	pylori or for use in detecting the presence of Helicobacter species in
CC	a sample. Antisense nucleic acid sequences of these sequences are



used to inhibit expression of a gene from *Helicobacter* species. *H.*  
*pylori* whole genomic DNA was isolated and nebulised to a median size of  
2000 bp. Purified DNA fragments were blunt-ended and ligated to unique  
BstXI linker adapters in 100–1000 fold molar excess. These linkers are  
complementary to the BstXI-cut PMX vectors, while the overhang is not  
self-complementary. Therefore the linkers will not concatemerise nor  
will the cut vector re-ligate itself easily. The linker-adaptor inserts  
were ligated to each of the 20 PMX vectors to construct a series of  
shotgun subclone libraries. The purified DNA samples were then  
sequenced.  
Note: The ORF/protein reference number for this sequence was obtained  
from the related specification, W09640893.  
Sequence 286 AA;

Query Match	15.28;	Score 227;	DB 22;	Length 286;
Best Local Similarity	28.68;	Pred. No. 3.65e+08;		
Matches	44;	Conservative	40;	Mismatches 62;
			Indels 8;	Gaps 7;

```
Dd      72 ekkdlnyigllkd-gepytsf-elknlgaf]wvlgelghematna-sla!slaldelnlee 128
          : ||| : - | : : : || - | : || : : : || : || : :
QY      3 DSDDIYAQLFQITDKGTAVDVYVDGEFYDHLSPQYGDDHTVLNALVAIAISYLEKIDVTN 62
```

[illegible]

```
Db 189 qahkysrlmdnleefkxflehcdrlilpyysa 222
   |:|:|:| |:|:|:|
QY 119 QPHTFSRTQAFLENEFAES-LSKADRVFLCEIFGS 151
```

RESULT	7
ID	W20102 standard; Protein; 286 AA

DT 29-JUL-1997 (first entry)  
DE H. pylori cytoplasmic protein, 11253.aa  
KW Cytoplasmic; vaccine; prevention; treat

KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis,  
 KW diagnosis.  
 OS Helicobacter pylori.

	key	location/qualifiers
FH		
FT	misc_difference	2
FT		/label= Unknown
FT		/note= "encoded by

FT	/label= Unknown
FT	/note= "encoded by TYA"
FT	misc_difference 166

FT	/note="encoded by RCC"
PN	W09640893-A1.
PD	19-DEC-1996.

07-JUN-1995; US-487032.  
01-APR-1996; US-630405.  
(ASTR ) ASTRA AB.

DR Helicobacter pylori nucleic acid sequen  
DR WPI; 97-052306/05.  
DR N-PSDB; T67723.  
PT Helicobacter pylori nucleic acid sequen

This sequence represents a H<sub>2</sub> nucleotide sequence of the H<sub>2</sub> gene, which is useful for vaccine development to detect Helicobacter infection, and to detect Helicobacter Claim 61, Page 335, 1481pp; English.

CC The protein may be used in a vaccine to  
CC infection or to identify *H. pylori* poly  
CC useful as potential *H. pylori* anti

The genomic sequence of *H. pylori* (ATCC 49239) overlapping contigs generated by mechanical shearing of DNA. The sequences were analysed for ORFs.

CC identify likely *H. pylori* antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from *H. pylori* by PCR amplification for recombinant polypeptide  
CC production, e.g. in *E. coli* hosts.  
SQ Sequence 286 AA;

Query Match	15.28;	Score 227;	DB 22;	Length 286;
...Best Local Similarity	28.68;	Pred. No. 3.65e-08;		
Matches	44;	Conservative	40;	Mismatches 62; Indels 8; Gaps 7,

```
Db 72 ekkdyniqylkcd-gepytsf-elknagafvlgigematna-slaissdelnhee 128
      : ||| : : : : : ||| : : : : : : : : : : : : : : :
QY 3 DSDDIYAQIGITDKGAVDVYVDGEFYDHFLSPOYGDHNYVNLALAVIAISLEKLDVTN 62
```

[illegible]

```
D8      189 qahkysrlmndleefkcflenhcdriilipvysa 222  
       |:|:| |::| :||  
QY     119 QPHFEFSRTQAFLNEFAES-LSKADRVFCEIFGS 151
```

RESULT 8  
ID W55117 standard: Protein; 335 AA.

DT	02-OCF-1998 (first entry)
DE	<i>Streptococcus pneumoniae</i> SP0067 protein.
KW	<i>Streptococcus pneumoniae</i> ; antiken; vaccine; infection; diagnosis;

OS	Streptococcus pneumoniae.
FH	key
FT	Misc,difference 1

PN	WO9818930-A2.
PD	07-MAY-1998.
PF	30-OCT-1997; U19422.

PA (HUMA-) HUMAN GENOME SCI INC.  
PI Chol GH, Hromockyj A, Johnson LS, Kunsch CA,  
DR WPI: 98-272224/24.

PT Nucleic acid encoding antigenic peptide(s) from *Streptococcus pneumoniae* - or their epitope-containing fragments, useful in protective or therapeutic vaccines, and for diagnosis

CC The present sequence represents a protein from *Streptococcus pneumoniae*.  
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
CC can be useful in vaccines for inducing protective antibodies against

Streptococcus pneumoniae, for treatment or prevention of infection e.g. pneumonia, otitis media or meningitis. Probes based on the nucleic acid are used to detect Streptococcus infection (by usual hybridisation or amplification methods) also for isolating Streptococcus genome or their

antibodies in standard immunoassays, especially for diagnosing or monitoring infections. Antibodies which bind the protein are used to detect corresponding antigens to confirm the infection and for treatment.

CC immunisation (optionally coupled to a toxin). Vaccines are administered  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
CC (especially 10-300)  $\mu\text{g/ml}$  per dose.

Query Match	9.2%;	Score 137;	DB 33;	Length 335;
Best Local Similarity	53.8%;	Pred. No. 9.70e-02;		

```
Db      161 gshvnenalatiavak]rdvdnqtlketlsafgvkhl 199
```



RESULT 9  
ID W29454 standard; Protein: 450 AA.  
AC W29454;  
DT 06-MAR-1998 (first entry)  
DE Streptococcus pneumoniae Mord protein.  
KM Mord gene; antibacterial agent; drug screening;  
KW uridine-diphosphate-N-acetylmuramyl-L-alanyl-D-isoglutamate ligase.  
OS Streptococcus pneumoniae.  
PN US5681694-A.  
PD 28-OCT-1997.  
PF 18-JUN-1996: 665435.  
PR 18-JUN-1996: US-665435.  
PA (ELIT ) LILLY & CO ELI.  
PI Hoskins JA, Peery RB, Skatrud PL, Wu CE;  
DR WPI: 97-535046/49.  
DR N-PSDB: T89154.  
PT Streptococcus pneumoniae Mord protein - useful for antibacterial  
PS drug screening  
PS Claim 2: Columns 21-22; 13pp; English.  
CC The present sequence represents Streptococcus pneumoniae Mord protein  
CC (uridine-diphosphate-N-acetylmuramyl-L-alanyl-D-isoglutamate ligase).  
CC The Mord protein is useful in a method for identifying compounds  
CC that inhibit Streptococcus pneumoniae Mord activity. Compounds  
CC identified are potentially useful as antibacterial agents.  
SQ Sequence 450 AA;

Query Match 9.2%; Score 137; DB 26; Length 450;  
Best Local Similarity 53.8%; Pred. No. 9,70e-02;  
Matches 21; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

Db 276 gshvenalatlavaklrldvngtlkclsfvgvkhrl 314  
QY 39 GDHTVNLALAVIAISYLEKLDVTNIKEALTEFGVKRRF 77

RESULT 10  
ID W68551 standard; Protein: 450 AA.  
AC W68551;  
DT 25-JAN-1999 (first entry)  
DE S. pneumoniae Mord protein.  
KM Biosynthesis; recombinant; antibacterial; bacterial cell wall;  
KW uridine-diphosphate-N-acetylmuramyl-L-alanyl-D-isoglutamate ligase;  
KM drug design; resistance; microorganism.  
OS Streptococcus pneumoniae.  
PN US5834270-A.  
PD 10-NOV-1998.  
PF 14-APR-1997: 843309.  
PR 18-JUN-1996: US-665435.  
PR 14-APR-1997: US-843309.  
PA (ELIT ) LILLY & CO ELI.  
PI Hoskins JA, Peery RB, Skatrud PL, Wu CE;  
DR WPI: 99-008720/01.  
DR N-PSDB: V33970.  
PT Streptococcus pneumoniae mur D biosynthetic gene - which encodes  
PT uridine-diphosphate-N-acetylmuramyl-L-alanyl-D-isoglutamate ligase,  
PT useful for identifying antibacterial compounds  
PS Claim 1: Column 13-15; 14pp; English.  
CC This is the amino acid sequence of the Streptococcus pneumoniae  
CC murD biosynthetic protein. The murD gene encodes the enzyme  
CC uridine-diphosphate-N-acetylmuramyl-L-alanyl-D-isoglutamate ligase.  
CC The nucleic acid can be used for the production of recombinant Mord  
CC protein and the implementation of large scale screens to identify  
CC new antibacterial compounds targeted at the stem peptide biosynthetic  
CC pathway involved in the synthesis of the bacterial cell wall.  
CC Structural analysis of the Mord protein will enable structure-based  
CC drug design to develop novel compounds for the treatment of antibiotic  
CC resistant microorganisms.  
SQ Sequence 450 AA;

Query Match 9.2%; Score 137; DB 37; Length 450;  
Best Local Similarity 53.8%; Pred. No. 9,70e-02;  
Matches 21; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

Db 276 gshvenalatlavaklrldvngtlkclsfvgvkhrl 314  
QY 39 GDHTVNLALAVIAISYLEKLDVTNIKEALTEFGVKRRF 77

RESULT 11  
ID W24790 standard; Protein: 1786 AA.  
AC W24790;  
DT 08-OCT-1997 (first entry)  
DE P. falciparum liver stage antigen-3.  
KM Plasmodium falciparum, pre-erythrocyte; liver stage antigen; serum;  
KW prophylaxis; Thai strain; gene organisation; exon; intron; hydrophobic;  
KW glycosyl-phosphatidylinositol membrane anchoring sequence; antibody;  
OS Plasmodium falciparum.  
FH Key Location/Qualifiers  
FT region 223..278  
FT /note= "repeat region 1"  
FT region 279..818  
FT /note= "repeat region 2"  
FT region 1537..1576  
FT /note= "repeat region 3"  
PN MO9641877-A2.  
PD 27-DEC-1996.  
PF 12-JUN-1996: F00894.  
PR 13-JUN-1995: FR-007007.  
PA (INSP ) INST PASTEUR.  
PI Daubersies P, Drullhne P;  
DR WPI: 97-065464/06.  
DR N-PSDB: T78868.  
PT Plasmodium falciparum poly:peptide(s) and related nucleic acids -  
PT derived from the liver stage antigen-3, useful for malaria vaccine  
PT prodn. and diagnosis  
PS Claim 1: Fig 2A-I: 69pp; French.  
CC This sequence corresponds to a Plasmodium falciparum strain K1  
CC pre-erythrocytic liver stage antigen-3 (USA-3) protein. The encoding  
CC gene sequence was isolated by screening a P. falciparum strain T9/96  
CC library with serum from a missionary treated by prophylaxis (for strain  
CC T6/96 see FR9101286). Of 20 clones isolated, clone 7295 was used to  
CC screen a library generated from Thai strain K1. One clone contained a  
CC 6.85 kb insert including the genomic sequence T78867. The gene comprises  
CC a 1.8 kb region encoding 3 major blocks of tetrapeptide repeats  
CC (especially the amino acid sequence VEEV, VEEN, VERI, VAPV, VAPR, etc)  
CC and a 3' hydrophobic region corresponding to a glycosyl-phosphatidyl-  
CC inositol membrane anchoring sequence. The invention relates to new  
CC polypeptides of at least 10 amino acids derived from the USA-3 protein  
CC with the exception of the peptides W24791-4. The USA-3 peptides can be  
CC used to raise antibodies and as vaccines for immunotherapy of malaria.  
SQ Sequence 1786 AA;

Query Match 9.0%; Score 135; DB 23; Length 1786;  
Best Local Similarity 20.4%; Pred. No. 1.33e-01;  
Matches 33; Conservative 55; Mismatches 66; Indels 8; Gaps 7;

Db 855 vvtllienveeta-ssvtfnnllelgegentlndtleekleehenyasaalentgse 913  
QY 50 IATSYLEKLDVTNIKEALTEFGVKRRFMTTIANQVYDDVNH-HPRESATIDARKK 108  
Db 914 eekkevidyleevkeevattllevegaeeksantlt-eifenleenave-snenvaenl 971  
QY 109 YPRKEVVAVAFOPPTFFSRTOAFLEFEAESLSKADRVLCFEFGSIRRENTGALTITODLIDKI 168  
Db 972 e--kl-netvfnvldkveetveisgeslenmemdaffseel 1010  
QY 169 EGASTLINESIN-VLEQFDNAVVLFGAGDIOKLOKNAVYIDLK 209

RESULT 12  
ID W20436 standard; Protein: 187 AA.  
AC W20436;  
DT 14-JUL-1997 (first entry)  
DE H. pylori protein.  
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;



CC H. pylori in a sample a

CC H. pylori anti

RESULT 14

ID W55244 standard; Protein: 293 AA.

AC W55244;

DT 26-JUN-1998 (first entry)

DE H. pylori ORF 05ae02020orf58 protein.

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;

KW Identification; binding compound; bacteria; life cycle; activator;

KW Inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;

KW bacterium.

OS Helicobacter pylori.

PN W09J75044.AL

PD 09-OCT-1997.

PF 27-MAR-1997; U05223.

PR 06-DEC-1996; US-761318.

PR 29-MAR-1996; US-625811.

PR 02-APR-1996; US-758731.

PR 25-OCT-1996; US-736605.

PR 28-OCT-1996; US-738859.

PA (ASTR ) ASTRA AB.

PI Alm RA, Smith D;

DR WPI; 97-503122/46.

DR N-PSDB: V24653.

PT Helicobacter pylori nucleic acid sequences and encoded

PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori

PT infection and for diagnosis of H. pylori infection

PS Claim 14; Pages 487-489; 1145pp. English.

CC This sequence is a Helicobacter pylori protein of unspecified

CC function.

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors.

CC The DNA and probes derived from it may be used for the

CC identification of H. pylori in a sample, and the diagnosis of

CC H. pylori infection. Nucleic acid sequences complementary to the

CC DNA act as antisense sequences, and can be used to prevent the

CC translation of H. pylori mRNA. Antibodies against the protein can

CC be used in immunoassays to evaluate the abundance and distribution

CC of H. pylori-specific antigens. The genomic sequence of H. pylori

CC (ATCC 55679) was determined from overlapping contigs generated by

CC mechanically shearing the bacterial DNA. The sequences were

CC analysed for ORF of at least 180 nucleotides, and the predicted

CC coding regions defined by computer evaluation. To identify likely

CC H. pylori antigens for vaccine development, the amino acid

CC sequences predicted from various ORF were analysed for significant







**THIS PAGE BLANK (USPTO)**



(174)











CC INFORMATION FOR SEQ ID NO: 17:

SQ SEQUENCE 411 AA; 47663 MW; 877573 CN

Query Match	Score	DB	Length
6.28;	92;	2;	411;







```
CC APPLICANT : Bumstead, Janene M.  
CC APPLICANT : Vermeulen, Arno N.  
CC TITLE OF INVENTION : Coccidiosis poultry vaccine  
CC NUMBER OF SEQUENCES : 10  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Akzo No. 5614195el Patent Department  
CC STREET: 1300 Piccard Drive, Suite 206  
CC CITY: Rockville  
CC STATE: Maryland  
CC COUNTRY: U.S.A.  
CC ZIP: 20850  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC CC  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/464,164  
CC FILING DATE: June 2, 1995  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Gormley, Mary E.  
CC REGISTRATION NUMBER: 34,409  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (301) 258-5200  
CC INFORMATION FOR SEO ID NO.: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 456 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC CC  
Qy SEQUENCE 456 AA; 52270 MW; 1035762 CN;  
  
Query Match          5.9%; Score 88; DB 1; Length 456;  
Best Local Similarity 21.2%; Pred. No. 6.5le+01;  
Matches      21; Conservative     26; Mismatches    46; Indels   6; Gaps   6,  
  
Db       284 LAAMALYMGSKLTNETDIDLKIFOKMDKN-GDGOLDKOELMEGIVELMKLKGEDVSV 342  
        ||.||:|||||::||::||::||::||::||::||::||::||  
Qy       47 LAVTAISLE-KLVNTNKEALET-FGVGRKFENETIANOVIDYAHHPREISAIDTR 104  
        |         |||||||::|::|::|::|::|::|::|::|::|  
Qz       105 ARKKYPHREYAVAFQHPTFSTQAFLNFASFESLSKA-DR 142  
        |         |||||||::|::|::|::|::|::|::|::|::|  
  
RESULT  12           STANDARD;             PRT;    456 AA.  
XX      ID      US -08 -668 -416 -2  
XX      AC      xxxxxx  
XX      DT  
XX      DE  
XX      EE  
XX      FF
```

```

CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: Patent In Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/668,416
CC      FILING DATE:
CC      CLASSIFICATION:
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/464,164
CC      FILING DATE: June 2, 1995
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Gormley, Mary E.
CC      REGISTRATION NUMBER: 34,409
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (301) 258-5200
CC      INFORMATION FOR SEQ ID NO: 2:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 456 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 456 AA; 52270 MW; 1035762 CN;
SQ
Query Match          5.9%; Score 88; DB 2; Length 456;
Best Local Similarity 21.2%; Pred. No. 6,51e+01;
Matches 21; Conservative 26; Mismatches 46; Indels 6; Gaps 6;
Db 284 LAAALALYMSKRLTNEETDELNLFQKMDKN-GDGLDQQLMEGYVELKIKGEDVSV 342
Oy 47 LAVIAISYLE-KLVYTNIKALELT-FGCVKRRFNENETIANQVIYDVVAHHPREISATIDT 104
Db 343 LDKSAITEFEQVLEAVDFDKN-GFI-EYSFVYVANDR 379
Oy 105 ARKKYPRKEVAVFQPHTESRTQAFLEEFESLSKA-DR 142
RESULT 13
ID US-08-736-770-6 STANDARD; PRT; 592 AA.
XX      xxxxxx
XX      Sequence 6, Application US/08736770
XX      Sequence 6, Application US/08736770
XX      Patent No. 5871965
XX      GENERAL INFORMATION:
XX      APPLICANT: Bandman, Olga
XX      APPLICANT: Au-Young, Janice
XX      APPLICANT: Hillman, Jennifer L.
XX      TITLE OF INVENTION: NOVEL HUMAN GUANYLATE BINDING PROTEINS
XX      NUMBER OF SEQUENCES: 6
XX      CORRESPONDENCE ADDRESS:
XX      ADDRESSEE: Incyte Pharmaceuticals, Inc.
XX      STREET: 3174 Porter Drive
XX      CITY: Palo Alto
XX      STATE: CA
XX      COUNTRY: US
XX      ZIP: 94304
XX      COMPUTER READABLE FORM:
XX      MEDIUM TYPE: Diskette
XX      COMPUTER: IBM Compatible
XX      OPERATING SYSTEM: DOS
XX      SOFTWARE: FastSEO Version 1.5
XX      CURRENT APPLICATION DATA:
XX      APPLICATION NUMBER: US/08/736,770
XX      FILING DATE: Filed Herewith
XX      PRIOR APPLICATION DATA:
XX      APPLICATION NUMBER:
XX      FILING DATE:
XX      ATTORNEY/AGENT INFORMATION:
XX      NAME: Billings, Lucy J.

```



```

CC      REGISTRATION NUMBER: 36,749
CC      REFERENCE/DOCKET NUMBER: PF-0145 US
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: 415-855-0555
CC      TELEFAX: 415-845-4166
CC      INFORMATION FOR SEQ ID NO: 6:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 592 amino acids
CC      TYPE: amino acid
CC      STRANDEDNESS: single
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: peptide
CC      IMMEDIATE SOURCE:
CC      LIBRARY: GenBank
CC      CLONE: 183002
CC      SEQUENCE 592 AA; 67902 MW; 1708054 CN;

Dy      13 LIENNTGLMANPEALKITLSAITPPMVV-VAIVGLYRTGSLYNNKLAKKKGF 65
Dy      164 LIDRIKGASLINEDSINYLEQFDNAVVLFWGAGDIQKLNAYL-DKL-GMKRNF 215

RESULT  14
ID      US-08-477-451-3          STANDARD:          PRT;      3174 AA.
XX      AC      xxxxxx
XX      DT
XX      DE      Sequence 3, Application US/08477451
XX      CC      Sequence 3, Application US/08477451
XX      CC      Patent No. 5928865
XX      CC      GENERAL INFORMATION:
XX      CC      APPLICANT: Covacci, Antonello
XX      CC      TITLE OF INVENTION: Helicobacter Pylori CagI Region
XX      CC      NUMBER OF SEQUENCES: 46
XX      CC      CORRESPONDENCE ADDRESS:
XX      CC      ADDRESSEE: Chiron Corporation
XX      CC      STREET: 4560 Horton Street
XX      CC      CITY: Emeryville
XX      CC      STATE: CA
XX      CC      COUNTRY: USA
XX      CC      ZIP: 94608-2916
XX      CC      COMPUTER READABLE FORM:
XX      CC      MEDIUM TYPE: Floppy disk
XX      CC      COMPUTER: IBM PC compatible
XX      CC      OPERATING SYSTEM: PC-DOS/MS-DOS
XX      CC      SOFTWARE: PatentIn Release #1.0, Version #1.30
XX      CC      CURRENT APPLICATION DATA:
XX      CC      APPLICATION NUMBER: US/08/477,451
XX      CC      FILING DATE: 07-JUN-1995
XX      CC      CLASSIFICATION: 435
XX      CC      ATTORNEY/AGENT INFORMATION:
XX      CC      NAME: McClung, Barbara G.
XX      CC      REGISTRATION NUMBER: 33,113
XX      CC      REFERENCE/DOCKET NUMBER: 0335.002
XX      CC      TELECOMMUNICATION INFORMATION:
XX      CC      TELEPHONE: 510-601-2708
XX      CC      TELEFAX: 510-655-3542
XX      CC      INFORMATION FOR SEQ ID NO: 3:
XX      CC      SEQUENCE CHARACTERISTICS:
XX      CC      LENGTH: 3174 amino acids
XX      CC      TYPE: amino acid
XX      CC      STRANDEDNESS: single
XX      CC      TOPOLOGY: linear
XX      CC      MOLECULE TYPE: protein
XX      CC      SEQUENCE 3174 AA; 369950 MW; 52965964 CN;

```

```

Query Match          5.9%: Score 88; DB 2; Length 3174;
Best Local Similarity 24.2%: Pred. No. 6.5le+01;
Matches 57; Conservative 37; Mismatches 72; Indels 7; Gaps 7

Db      2993 LTRIPKFLQSKAKQEKMLNAKH-SNSLVYSGCHGRWIGSSLLSLALGRF-YLLLVA 3050
        :|::||::||::||::||::||::||::||::||::||::||::||::||::||
Oy      58 LDVNIKEALETFPGVGVRREMETITIANQVIYDDVAHHPRELSATIDARK-KYHKREYA 116
        ::||::||::||::||::||::||::||::||::||::||::||::||::||
Oy      117 VFOPHTFSRTQAFINFEAFESLSKDRAFLCEIFGISIRENGALTIOPLDIKEGASLINE 176
        :|::||::||::||::||::||::||::||::||::||::||::||::||
Db      3108 ALPYRHNRFHOIALLKGFLK-RLPHRYOKRL 3139
        ::||::||::||::||::||::||::||::||::||::||::||::||
Oy      177 DSNVLEQFDNAVVLFWMGAGDIOQLQNAYLDKL 209
        ::||::||::||::||::||::||::||::||::||::||::||::||

RESULT   15
ID       US-08-471-119A-2         STANDARD:      PRT:    15281 AA.
AC       xxxxxx
DT
DX
Sequence 2, Application US/084711119A
DE
XX
Sequence 2, Application US/084711119A
CC
Patent No. 5827706
CC
GENERAL INFORMATION:
CC
APPLICANT: Leitner, Ernst
CC
APPLICANT: Schneider, Elisabeth
CC
APPLICANT: Schoengenderfer, Kurt
CC
APPLICANT: Weber, Gerhard
CC
TITLE OF INVENTION: Cyclosporin Synthetase
CC
NUMBER OF SEQUENCES: 8
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSEE: NO. 5827706artis Corporation
CC
STREET: 59 Route 10
CC
CITY: East Hanover
CC
STATE: New Jersey
CC
COUNTRY: USA
CC
ZIP: 07936
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: Floppy disk
CC
COMPUTER: IBM PC compatible
CC
OPERATING SYSTEM: PC-DOS/MS-DOS
CC
SOFTWARE: PatentIn Release #1.0, Version #1.25
CC
CURRENT APPLICATION DATA:
CC
APPLICATION NUMBER: US/08/471,119A
CC
FILING DATE: 06-JUN-1995
CC
CLASSIFICATION: 435
CC
ATTORNEY/AGENT INFORMATION:
CC
NAME: Kassenooff, Melvyn
CC
REGISTRATION NUMBER: 26,389
CC
REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
CC
TELECOMMUNICATION INFORMATION:
CC
TELEPHONE: 201 503 8474
CC
TELEFAX: 201 503 8807
CC
INFORMATION FOR SEO ID NO: 2:
CC
SEQUENCE CHARACTERISTICS:
CC
LENGTH: 15281 amino acids
CC
TYPE: amino acid
CC
STRANDEDNESS: single
CC
TOPOLOGY: unknown
CC
MOLECULE TYPE: protein
CC
HYPOTHEICAL: NO
CC
ANTI-SENSE: NO
CC
ORIGINAL SOURCE:
CC
ORGANISM: Tolypocladium niveum
CC
STRAIN: ATCC 34921
S0     SEQUENCE 15281 AA; 1689051 MW; 1163257458 CN;

Query Match          5.9%: Score 88; DB 2; Length 15281;

```



Mon Nov 29 10:17:02 1999

US-09-103-287-4.rai

Page 8

Best Local	Similarity	24.18;	Pred. No.	6.51e+01;					
Matches	20;	Conservative	22;	Mismatches	36;	Indels	5;	Gaps	5;

Matches	20;	Conservative	22;	Mismatches	36;	Indels	5;	Gaps	5;
---------	-----	--------------	-----	------------	-----	--------	----	------	----

Db 12604 VFEADLFEATIRSVVEVHEILRRGLDQPIAISTMPLVDGLAALNSRNLPAVEDIEP 12663

QY 52 ISYLEKL-DVINIKEALETFGGVKRR-FNETTIA-NQV-IVDDYAH-HPREISATIDTAR 106

Db 12664 DFATEASVDVEQTQVNPDAL 12686

1111 : :1:  
QY 107 KKYRKEVAVQPHTFSTRQAF 129

Search completed: Wed Nov 24 00:54:20 1999  
Job time : 47 secs.

Job time : 47 secs.



\*\*\*\*\*  
 WISE (TM)  
 \*\*\*\*\*

Release 3.1a John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Mpsrch\_tpn n.a. - n.a. Smith-Waterman search, using a protein query  
 which has been backtranslated into n.a. using IUPAC symbols

Run on: Sat Nov 27 09:38:23 1999; MasPar time 1198.35 Seconds  
 1491.947 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-4  
 Description: (1-215) from US09103287.pep  
 Perfect Score: 2663  
 N.A. Sequence: 1 TTVAARGAYMSNGAAGAYAT.....TNGCNATGAAARACNTTY 645  
 Comp: AARTTYCTRMSCNCTRCRTA.....ANCNTACTTTRCGNAR

Scoring table: TABLE dktranslat2  
 Gap 30

Mmatch SMD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953653 bases x 2

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: embd158  
 1:em\_ba1 2:em\_ba2 3:em\_fun 4:em\_htg 5:em\_hum1 6:em\_hum2  
 7:em\_in 8:em\_om 9:em\_ov 10:em\_ov 11:em\_pat 12:em\_ph  
 13:em\_pl 14:em\_ro 15:em\_sts 16:em\_vl

Database: genbankx11  
 17:gb\_ba1 18:gb\_ba2 19:gb\_htg1 20:gb\_htg2 21:gb\_in1  
 22:gb\_in2 23:gb\_om 24:gb\_ov 25:gb\_pat 26:gb\_ph 27:gb\_pl1  
 28:gb\_pl2 29:gb\_pl1 30:gb\_pl2 31:gb\_pl3 32:gb\_ro  
 33:gb\_st 34:gb\_sts 35:gb\_sy 36:gb\_un 37:gb\_vl

Statistics: Mean 65.249; Variance 231.556; scale 0.282

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	2573	96.6	1314	17	AF034076	8.76e-375
2	1309	49.2	2320	17	BACINAM	3.43e-125
3	1309	49.2	2136	17	BSUB0016	3.43e-125
4	1309	49.2	2200	18	AF008220	3.43e-125
5	324	12.2	1483	18	AE001213	4.01e-15
6	320	12.0	2454	17	POYMURC	1.00e-14
7	318	11.9	15079	18	AE001180	1.58e-14
8	286	10.7	15862	18	AE000736	2.11e-11
9	283	10.6	1998	17	EC067892	4.11e-11
10	283	10.6	21757	18	ECMORGC	4.11e-11
11	283	10.6	21757	18	AE000118	4.11e-11
12	283	10.6	28377	17	EC2MIN	4.11e-11

13	283	10.6	111401	17	EC0110K	4.11e-11
14	283	10.6	194271	19	AC005903	4.11e-11
15	279	10.5	312430	17	SPX02	9.54e-11
16	277	10.4	22268	17	SPB86147	1.54e-10
17	275	10.3	12085	18	U32794	2.40e-10
18	268	10.1	7218	25	166494	1.11e-09
19	247	9.3	6757	18	AF012886	1.02e-07
20	234	8.8	11681	18	AE001348	1.58e-06
21	224	8.4	148142	17	D90916	1.26e-05
22	222	8.3	14577	18	AE001489	1.90e-05
23	213	8.0	12291	17	AB015023	1.19e-04
24	198	7.4	13126	18	HPAE000576	2.35e-03
25	195	7.3	16157	18	AE001670	4.22e-03
26	183	6.9	1350	25	170976	4.24e-02
27	183	6.9	1350	25	170975	4.24e-02
28	173	6.6	161453	31	AC004993	1.90e-01
29	173	6.5	8000	18	AF068902	2.75e-01
30	168	6.3	37586	17	MTCY270	6.85e-01
31	162	6.1	38793	17	MSCB27CS	2.01e+00
32	163	6.1	216021	31	HUAC004787	1.69e+00
33	157	5.9	202879	31	HUAC005859	4.86e+00
34	156	5.9	220000	20	AC004795	5.78e+00
35	154	5.8	2644	17	SSAHCYGN	8.19e+00
36	154	5.8	34397	21	CEP47G4	8.19e+00
37	154	5.8	110000	19	CEP63D3	8.19e+00
38	153	5.7	1934	17	BASIGEP	9.73e+00
39	153	5.7	3381	24	CHKREBPA	9.73e+00
40	153	5.7	236120	37	AF063866	9.73e+00
41	148	5.6	316	28	IMAXITS1	1.99e+01
42	149	5.6	2126	21	MOSAWB	2.28e+01
43	148	5.6	2447	32	MUSGERS	2.28e+01
44	149	5.6	23112	17	AB010415	1.93e+01
45	150	5.6	74371	31	AC003369	1.63e+01

## ALIGNMENTS

RESULT	1	AF034076	1314 bp	DNA	BCT	26-NOV-1997
LOCUS		Staphylococcus aureus UDP-N-acetylmutramoyl-L-alanine synthetase				
DEFINITION		(murC) gene, complete cds.				
ACCESSION		AF034076				
NID		g2642658				
VERSION		AF034076.1	GI:2642658			
KEYWORDS						
SOURCE						
ORGANISM		Staphylococcus aureus.				
REFERENCE		Staphylococcus aureus				
AUTHORS		Eubacteriae; Firmicutes; Low G+C gram-positive bacteria; Bacillaceae; Staphylococcus.				
TITLE		1 (bases 1 to 1314)				
JOURNAL		Lowe, A.M. and Deresiewicz, R.L.				
AUTHORS		Cloning and sequencing of Staphylococcus aureus murC, a gene involved in cell wall biosynthesis				
TITLE		Unpublished				
JOURNAL		2 (bases 1 to 1314)				
AUTHORS		Lowe, A.M. and Deresiewicz, R.L.				
TITLE		Direct Submission				
JOURNAL		Submitted (11-NOV-1997) Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Ave, Boston, MA 02115, USA				
FEATURES						
source		location/Qualifiers				
		1. 1314				
		/organism="Staphylococcus aureus"				
		/db_xref="taxon:1280"				
	gene	1. 1314				
		/gene="murC"				
		1. 1314				
		/gene="murC"				
		/function="cell wall biosynthesis"				
		/note="MurC: UDP-N-acetylmuramate-alanine ligase"				
		/codon_start=1				
		/transl_table=1				
		/product="UDP-N-acetylmuramoyl-L-alanine synthetase"				



		/protein_id="AAB87090.1"	
		/db_xref="pid:g2642659"	
		/translation="MTHTHTVGIKSGMSLSLAQIMHDLGHEVGSQDIENYFTEVALR NKIKILIPDANNIKEDMVIQGNARSSHEERAKRQMLDVSIIDELGQIIDQYT SVAVTGAHGTSTTGLSHVNDGDKTSFLIGDTGQMLSEYFAEACEYRHFUS YKPDYAIMTIDIDPHYEDINDVPEDEMAHMYKGIANGDEHLEKRIADYPI YXXFKSDSDIYAKQNIITDKGTAFDYYVGEYDHFSPQGDHVTALAVIAISY LEKLDVNIKEALETGGVARRFNETTIANOVIVDVAHHPRETSIETARKKYPK EYAVFOPHFSRFOAFLENEFASISADDFCEIFGSIRENTGATIIDLDIKIEG ASLINEDSTVLEQFDNAVYLFAGAGCIGTQLQNAIYDKLKMNAF"	
BASE COUNT	467 a 163 c 250 g 434 t		
ORIGIN			
Query Match	96.6%; Score 2573; DB 17; Length 1314;		
Best Local Similarity	59.5%; Pred. No. 8.76e-275;		
Matches	384; Conservative 158; Mismatches 103; Indels 0; Gaps 0;		
Db	667 TTAAAGATTCGATGACATTTATGCTCAAAATATTCAAATTAAGATTAAGTACTGCT 726		
Oy	1 TTAAAGATTCGATGACATTTATGCTCAAAATATTCAAATTAAGATTAAGTACTGCT 726		
Db	727 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 786		
Oy	61 GTNAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120		
Db	787 CATACATTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 846		
Oy	121 CAYCAGTNTNTNATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180		
Db	847 ACAATATTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 906		
Oy	181 ACNAATATTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240		
Db	907 ACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 966		
Oy	241 ACNATHTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300		
Db	967 ACAATTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1026		
Oy	301 ACNATHTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360		
Db	1027 CACACTTTCTAGACACAGCATTTTAAATGATGATGATGATGATGATGATGATGATGAT 1086		
Oy	361 CAYCAGTNTNTNATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420		
Db	1087 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1146		
Oy	421 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480		
Db	1147 ATACAGATTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1206		
Oy	481 ATACAGATTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540		
Db	1207 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1266		
Oy	541 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600		
Db	1267 TTACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1311		
Oy	601 TTACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 645		
RESULT	2		
LOCUS	BACUNAM 2320 bp DNA BCT 26-MAR-1997		
DEFINITION	Bacillus subtilis UDP-N-acetylmuramate-alanine ligase gene, partial		
ACCESSION	L31845		
NID	9556013		
VERSION	L31845.1 GI:556013		
KEYWORDS	UDP-N-acetylmuramate-L-alanine ligase.		
SOURCE	Bacillus subtilis.		
ORGANISM	Bacillus subtilis		
REFERENCE	Eubacteria; Firmicutes; Low G+C gram-positive bacteria; Bacillaceae; Bacillus.		
AUTHORS	Varon, D., Brody, M.S. and Price, C.W.		
TITLE	Bacillus subtilis operon under the dual control of the general stress transcription factor sigma B and the sporulation transcription factor sigma H		
JOURNAL	Mol. Microbiol. 20 (2), 339-350 (1996)		
MEDLINE	96310371		
FEATURES			
source	location/Qualifiers		
	1..2320		
	/organism="Bacillus subtilis"		
	/strain="168"		
	/sub_species="Marburg"		
	/db_xref="taxon:1423"		
	<1..1049		
	/note="Identification based on sequence similarity to E. coli enzyme"		
	/codon_start=3		
	/transl_table=11		
	/product="UDP-N-acetyl muramate-alanine ligase"		
	/protein_id="AAB40043.1"		
	/db_xref="pid:g556014"		
	/db_xref="GI:556014"		
	/translation="IPVIRYHKEFLGDYMKKTSVAVTGAHGTSTGLAHVIONAKP TSFLIGDTGOGENSEYFVFEACEYRHLISQPYATMTNDPDPYESSIDPVE DAFOEMALOVNKGITACGDDEHLPKHAVAPVYVCTGEOQANIKYKSTGCTDVE VPRNFYPTFTYPRAYGHNVLSLAVIALCHEEDDSITIKALKSFGCVKRFREK QUGDOVLIDYDHPHETVITVLEAOKRPDEIVAVFQPHFTTRQDFLDEFAELS GADCVYLCDFISGARENAGKLTIGLOGKIHNAKLEEDTSVLKADRAVLIEMGAG DIQKYRAVENVMA"		
	1048..1079		
	/note="sigma-H consensus"		
	1101..1109		
	/note="sigma-H consensus"		
	1123..1127		
	/note="sigma-H consensus"		
	prim_transcript 1136		
	/note="sigma-H dependent 5' end"		
	-35_signal 1151..1158		
	/note="sigma-B consensus"		
	-10_signal 1172..1177		
	/note="sigma-B consensus"		
	prim_transcript 1186		
	/note="sigma-B dependent 5' end"		
	1198..1205		
	/note="ORF1"		
	1211..1642		
	/codon_start=1		
	/transl_table=11		
	/protein_id="AAB40044.1"		
	/db_xref="pid:g556015"		
	/db_xref="GI:556015"		
	/translation="MTILYSLVALIAFVLVLYLSLTKSLQTLTKNASTLEGL EQMKGITETAELELHKTNRLEAIEDOEKSEKNTVVAQGVASVOQENTSMKQAG SVASVRENODKINQVGRSTWSQAAMEIWEKKQKKSAL"		
	1660..1665		
	/note="similar to plant water stress proteins; ORF2"		
	1670..2128		
	/codon_start=1		
	/transl_table=11		
	/protein_id="AAB40045.1"		
	/db_xref="pid:g556016"		
	/db_xref="GI:556016"		
	/translation="MKSKDINSKDFLTIGIGITGATTAFLAPSKSEKRLDGLS QAVALLDKTKMTADAKERGTQYVSIANKTSTNTQVLADSGQIMNKYKDLDRSKS DKTSDSTAMQDREEMQAADETKQVLQTKEDYKDELKDAQKQAEQLNR"		
	2135..2139		
	/note="open reading frame extends an additional 53 codons (including the TGA stop) into the overlapping sequence reported by Bolotin et"		
	/codon_start=1		
	/transl_table=11		



DOATTSB/K./ DOATSTER/D./ DIANS/D./ DIANN/D./ DIAGNET/D.C./

```

        . /product="hippurate hydrolase"
        /protein_id="CAB14907.1"
    
```

*Bacillus subtilis*.  
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;  
1 (bases 1 to 215640)  
Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,  
Azenedo,V., Bertero,M.G., Bessières,P., Bolotin,A., Borcherdt,S.,  
Doriss,R., Bourcier,L., Evans,A., Braun,M., Bridgell,S.C.,  
BSUB0016 215640 bp DNA BCT 26-NOV-1997  
*Bacillus subtilis* complete genome (section 16 of 21): from 2997777.  
to 3213410. 299119 AL009126  
G2635411  
Z99119.1 GI:2635411



gene  
/db\_xref="PID: e1185803"  
/db\_xref="PID: g2635413"  
/db\_xref="GI: 2635413"  
/translacion="MSLDYWRNI EGSYPQTGNDILTLKEESNPNVLSLEKOLIGI  
RRLHLYPPLSKSEEFETPAFIKKCKLEKGIQIRPALTKGVADIAJGSEBPALALRR  
DIDALDEETKGTPYPAKSKHKGIMHACGHDHFHAALIGAAPFLKEMODSKIKILLR  
PAEENAGATKYTIEDGQLDGIDNIVGLHNKPIANGTYGLKTGPILMAVDFPKYIEES  
KGNALVPNGDEPILIGASQLI VALQTVSRNVP LQSAIILTVGKINGSGSWNIIPDT  
VIGELVRFPDESRNQVOKREFAVTEQISAFAFSKNAVKNMGPPLCNDPEATGLV  
ROAHNKAQLQVIDPASTAGEDFAVYLEHIPGSFAFFETGCDGHDHNAFFIIDEFAIIT  
KASYFAKESAKRLDNEESKISD"  
complement(2288. 2980)  
/gene="ribR"  
complement(2288. 2980)  
/gene="ribR"  
/function="regulation of riboflavin biosynthesis genes"  
/note="alternate gene name: ytnk"  
/codon\_start=-1  
/transl\_table=11  
/product="transcriptional regulator"  
/protein\_id="CAB14908.1"  
/db\_xref="PID: e1185803"  
/db\_xref="PID: g2635414"  
/db\_xref="GI: 2635414"  
/db\_xref="SWISS-PROT: P94465"  
/translacion="MTIILGVYKGRKQGRKLGFPANVDAKIHGLRNGVYGLATVNA  
HOFHGLVNNYKGRPTVSGNLKLETLFI LFDHFDIYIGKIECSILFKIRERRPDSLE  
SLTKQKKNICSVAKRPELIGIMANKKESLSHOLNLPDLGCFKKCNNTYVNRGAG  
VYVIDMFEYGIQTVAVYRRIYIILSFPLKEDNPKVSKYIRGAGSLAKLNRFIS  
SYVEEENILIS"  
complement(3027. 4355)  
/gene="ytnJ"  
complement(3027. 4355)  
/gene="ytnJ"  
/function="unknown"  
/note="similar to htrIIlriacetate monooxygenase"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB14909.1"  
/db\_xref="PID: e1185804"  
/db\_xref="PID: g2635415"  
/db\_xref="GI: 2635415"  
/db\_xref="SPRMBL: O34974"  
/translacion="MTRADFIQFGAMIHGVGGTTGWRHPVDPSASTNIEFYKKAQ  
TAELKLFSTFIADLFISEKSIIPHRNFPEPITLSASVTKNIGLVGFSSPFE  
PTTIRLOLSLHISGRAGMNLVTSPOEJAARNRKNLPEHTEYELIOEHLDVVR  
GIJMSNEHDAFIHNKTKGQFPDOAKIHLINKKRGKFFQVEGELNIGRKSQSPVYQAG  
SSETGRQPAKRNADAIFTHSNSLETKAFIYADVSKRADBEGRDSSVSIIPFGISPIVA  
DIEEERKEAFREFALPIENAVYTLAEFPDYDLSVPLDEPFDIGDVKNAFOST  
TDRIRKAKARNULRLREVAQEMAFPRTLFI GTPERVASLIETWFNAEADDFI VSGDI  
POTLDAFEKVYPIQERGLIYKODYRGGLTRENIGLGIPIHQSVLSHSHH"  
complement(4352. 4633)  
/gene="ytnI"  
complement(4352. 4633)  
/gene="ytnI"  
/function="unknown"  
/codon\_start=-1  
/transl\_table=11  
/protein\_id="CAB14910.1"  
/db\_xref="PID: e1185805"  
/db\_xref="PID: g2635416"  
/db\_xref="GI: 2635416"  
/db\_xref="SPRMBL: O34639"  
/translacion="MSDVNIYVMSKSCSYCEFYKNLKNKSGPPONIOVSEKELR  
DILQYKGYRHPVYVEIRGNQYOSITETIGIEHDDLALNHAQIKEAKR"  
complement(4648. 5652)  
/gene="ytnM"  
complement(4648. 5652)  
/gene="ytnM"  
/function="unknown"  
/note="similar to hypothetical proteins"  
/codon\_start=-1

gene  
CDS  
gene

```

/transl_table=11
/protein_id="CAB14911.1"
/db_xref="PID:e1185806"
/db_xref="PID:g2635417"
/db_xref="GI:2635417"
/db_xref="SPTREMBL:O34846"
/transl_table="MIRSLIDSLGGGTAADTLQHTVKAQMAEEGCHRRHWE
HNNNDKSGAPSVLYGLCYLAATSKRIKRLASGCVLQWYSSKYAEQHLLSALPGI
DLGVKAKGCGQLSDLAQAEIKKPVROFDEKLELTHFYRDDPDTHTAALPRQ
VDRKGIQLLGGSTESAISAALKIGSEFAFVINGEEVLEKEARAPADHALPGSESE
FLHAAVVAANTKEAEKHIVRSRKIVKLVDGRKVVNGSREQAEALEVNTPPYDI
VQKTIINGTKEEVAEELRLSGTYKINDVIFTPRIKNAVEKOLSYQLSDAVLAAR
"
-complement(5649. .6428)
/gene="hisp"
complement(5649. .6428)
/gene="hisp"
/codon_start=1
/transl_table=11
/product="histidine transport protein (ATP-binding
protein)"
/protein_id="CAB14912.1"
/db_xref="PID:e1185807"
/db_xref="PID:g2635418"
/db_xref="PID:g2635418"
/db_xref="GI:2635418"
/db_xref="SPTREMBL:O34900"
/transl_table="MIEIKINIKQFGIHHVLCGINTLVKRGEEVYIIIPSGSKTTEL
RCNLNLEPRDDGIISHDKVINCRPSPSKREYHMLRKQTAVFQOYHLFAKTVIENW
EGLTFAARKRQDGAVAANENLRKVGQDKLNAYPSQSGGKORVGIARALAHDDV
LLFDPTAALPDELVEGLVEGLVEGLVEGLVEGLVEGLVEGLVEGLVEGLVEGLVE
EGGTPEEVRFRKTKDRTRQPIKRVSPPELPEPKHKEPVI"
complement(6425. .7132)
/gene="ytum"
complement(6425. .7132)
/gene="ytum"
/function="unknown"
/note="similar to amino acid ABC transporter (permease)"
/codon_start=1
/transl_table=11

```

Note: remainder of annotations omitted.

Query Match	49.28;	Score 1309;	DB 17;	Length 215640;
Best Local Similarity	45.28;	Pred. No. 3.43e-125;		
Matches	272;	Conservative 123;	Mismatches 207;	Indels 0;
				Gaps 0.

Db	49508	GTGAGCGTCATATAATTTTGGATATCTCCGTCGCCATGAAGATGGAGCGCTTATC	49567
Cp	615	RIANGRRTTYTGNARTTYYTGDATCTCCGCGCCCATRAAARNAACNNCCRTTRTC	556
Db	49568	ATAGAGCTTTTAAACAGATGTCTACTCTTCATCAGCGTGGCATTTATGAATTTTCC	49627
Cp	555	RAAYTCCTCNARACTTDDATSMRTCTCTCTTDTAARNSMNCNCCYCDATYTTTRC	496
Db	49628	CTGAGATACCGATCGTCAGCTTTCGGCATCTCTCAGCGGCTGAGCGGAATATTCGA	49687
Cp	495	DATNARTCTYGDATNGTANRAGCCNCGTRTTYTCNCKDATNSMNCRAADATYTCRA	436
Db	49688	TAATATCACACAGTGGCGGCGCCCTCAGGCTTCTTCGAAATTCGCAAGAGACGCTGCCT	49747
Cp	435	NARRAANACNCRKTCAGCTTNSMARRSWTCCGACRAATCTRTTAARRAANCCYTGNET	376
Db	49748	CGCGGTAAATGTATGAGCGTGGAAATCCGACACATTTCCGATCAGAGATCTGTCT	49807
Cp	375	NCKNSMRANNGRTNGGTYTGRAAACAGCAACACCTCTTTRGNGRATYTTTTRNC	316
Db	49808	TGCCGCGCTGATTGTCACTTTTATTTCTGCGATGATGGGCAATGATCAATCAGCAC	49867
Cp	315	NGCNGRTCDANNGTNGCNSMDATYTCNCKNGSGRTGNGCARTCRTCNACDATTNAC	256
Db	49868	TTGATATCCCAAGCTGCTTCTCATTTGAATCTCGTTTACGCCCCCAAGATTGGAGAC	49927
Cp	255	YTRRTTNGDANNGTNGYTRCTTTAAACNCKKCTYTTAACNCCNCCRAADYTYCNARNC	196







AUTHORS Yocum,R.R., Perkins,J.B., Howitt,C.L. and Pero,J.  
TITLE Cloning and characterization of the metE gene encoding  
S-adenosylmethionine synthetase from *Bacillus subtilis*  
JOURNAL J. Bacteriol. 178 (15), 4604-4610 (1996)  
MEDLINE 96345628  
REFERENCE 21 (bases 100760 to 102298)  
AUTHORS Kappes,R.M., Kempf,B. and Bremer,E.  
TITLE Three transport systems for the osmoprotectant glycine betaine  
operate in *Bacillus subtilis*: characterization of *Opd*  
JOURNAL J. Bacteriol. 178 (17), 5071-5079 (1996)  
MEDLINE 96359357  
REFERENCE 22 (bases 1 to 220060)  
AUTHORS Lapidus,A., Galleron,N., Sorokin,A. and Ehrlich,S.D.  
TITLE Sequencing and functional annotation of the *Bacillus subtilis* genes  
in the 200 kb *rmb-dnaB* region  
JOURNAL Microbiology 143 (Pt 11), 3431-3441 (1997)  
MEDLINE 98048467  
REFERENCE 23 (bases 1 to 220060)  
AUTHORS Lapidus,A., Galleron,N., Sorokin,A. and Ehrlich,S.D.  
TITLE Direct Submission  
JOURNAL Submitted (13-JUN-1997) Laboratoire de Genetique Microbienne, INRA,  
Domaine de Vilvert, Jouy-en-Josas cedex 78352, France  
FEATURES  
Source Location/Qualifiers  
1..220060  
/organism="Bacillus subtilis"  
/db\_xref="taxon:1423"  
485..2034  
/product="16S ribosomal RNA"  
2203..5129  
/product="23S ribosomal RNA"  
5185..5300  
/product="5S ribosomal RNA"  
5322..5397  
/note="codon recognized: GGA"  
/product="tRNA-Val"  
5430..5504  
/note="codon recognized: ACA"  
/product="tRNA-Thr"  
5543..5618  
/note="codon recognized: AAA"  
/product="tRNA-Lys"  
5629..5715  
/note="codon recognized: CTG"  
/product="tRNA-Leu"  
5721..5795  
/note="codon recognized: GGC"  
/product="tRNA-Gly"  
5810..5895  
/note="codon recognized: TTA"  
/product="tRNA-Leu"  
5905..5981  
/note="codon recognized: CGT"  
/product="tRNA-Arg"  
5997..6073  
/note="codon recognized: CCA"  
/product="tRNA-Pro"  
6079..6151  
/note="codon recognized: GCA"  
/product="tRNA-Ala"  
6172..6248  
/note="codon recognized: ATG"  
/product="tRNA-Met"  
6251..6327  
/note="codon recognized: ATG"  
/product="tRNA-Ile"  
6334..6425  
/note="codon recognized: TCA"  
/product="tRNA-Ser"  
6443..6519  
/note="codon recognized: ATG"  
/product="tRNA-Met"  
6531..6607  
/note="codon recognized: GAC"

Note: remainder of annotations omitted.

Query Match 49.2%; Score 1309; DB 18; Length 220060;  
Best Local Similarity 45.2%; Pred. No. 3,43e-125;  
Matches 272; Conservative 123; Mismatches 207; Indels 0; Gaps 0;

Db 130317 ATGATTTTCAAGCCGAAACATCGTAAACACACGAGGAGGCACTTTGATGCTTTG 130376  
14 AAGATATTTATGCGNCAARHTTTCARATHACNGAATAAGNAGNCNNTNGATGNTXG 73

Db 130377 TCCGCAATACGTTCTATGATACGTTTATTCCTGCTAGCGCCACCAATGATTAA 130436  
74 TNGAATGNGARTTAYAYACATTTTNTWMSGCCARTATGNGNAYCAVACNGNTYNA 133

Db 130437 ACTCATTTGGCGTCACTTTCGTTATGCCATTTATGAAGAAATGATTCGCAATTAAGC 130496  
134 AAGCMTTNGCGNTNATTHCCNATHTWSNTAYTNGAARATNGAATNCAATHAARG 193

Db 130497 ATGCTCTCAAAATCCTTTGGGGGCGTCAACGACAGTTCATGAGAGCAGCTTGGGATC 130556  
194 AAGCTTTGATGATGATTTGNGNGNNGTAAARMGNTTAAAGARACNCAATHGCNAVC 253

Db 130557 AAGTGTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 130616  
254 AAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 313

Db 130617 CAAGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 130676  
314 CAGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 373

Db 130677 GGACGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 130736  
374 GACGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 433

Db 130737 TATGCGATTTTGGGCTGACCGCGTGAAGATGCGGAGAGCTGATGATGATGATGATG 130796  
434 TATGCGATTTTGGGCTGACCGCGTGAAGATGCGGAGAGCTGATGATGATGATGATGATG 493

Db 130797 AGGGAATTCATTAATGCCAAGCTGATGAAAGATGACACATCTTTTAAAGCTC 130856  
494 THGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 553

Db 130857 ATGATTAAGCCGTTCTCATCTTCATGAGAGAGATGATGATGATGATGATGATGATG 130916  
554 TYGATTAAGCCGTTCTCATCTTCATGAGAGAGATGATGATGATGATGATGATGATGATG 613

Db 130917 AC 130918

QY 614 AY 615

RESULT 5  
LOCUS AE001213 14838 bp DNA BCT 16-JUL-1998  
DEFINITION Treponema pallidum section 29 of 87 of the complete genome.  
ACCESSION AE001213 AE000520  
NID G3322606  
VERSION AE001213.1 GI:3322606  
KEYWORDS  
SOURCE Treponema pallidum.  
ORGANISM Treponema pallidum  
Eubacteria; Spirochaetales; Spirochaetaceae; Treponema.  
REFERENCE  
AUTHORS Fraser,C.M., Norris,S.J., Weinstein,G.M., White,O., Sutton,G.G.,  
Dodson,R., Gwin,M., Hickey,E.K., Clayton,R., Ketchum,K.A.,  
Sodergren,E., Hardham,J.M., McLeod,M.P., Salzberg,S., Peterson,J.,  
Khalak,H., Richardson,D., Howell,J.K., Chidambaram,M.,  
Uitterlisch,T., McDonald,L., Artlich,P., Bowman,C., Cotton,M.D.,  
Fujii,C., Garland,S., Hatch,B., Horst,K., Roberts,K., Matthey,L.,  
Weidman,J., Smith,H.O. and Venter,D.C.  
TITLE Complete genome sequence of *Treponema pallidum*, the syphilis  
spirochete  
JOURNAL Science 281 (5375), 375-388 (1998)















```

...
Note: remainder of annotations omitted.

Query Match      11.9%   Score 318; DB 18; Length 15079;
Best Local Similarity 37.4%; Pred. No. 1,586-14;
Matches 76; Conservative 44; Mismatches 83; Indels 0; Gaps 0;

Db 4542 TGGAGATTATGCTCATCATCTCCTTAGGGAATAATAAATACCTCTTTTGGATTAAATTT 4601
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 263 TTNGVAGATTACGCACAYCACAYCCNNMGNGRARRHWMGNCNCAATFGAAYCNCNGGNARA 322
|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 4602 TTATTAACAATAAAGCGATATATTTTGATTTTGTATTCCTCATCCTTACAGAACAAAG 4661
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 323 ARTATGCCATTAAGAARGRTNGTNGCNGNTNTTYGCARCCCATACACTTTYSNMGNACNCRG 382
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 4662 AATTTTTTCGGATTTTGTGTAAGTTCTTAAGTGTCTGCGATATATTTATTTTGCACATA 4721
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 383 CNTTYTNAAGARTTYGCGNCFARMSNTYMENMARGCNAGVMGNGTWTYYTNTGYCARA 442
|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db 4722 TATATCTTTCAAATAGGGAAT 4744
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY 443 THTTYGGMNSMATHMGNARAY 465
|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||

RESULT      8          AE000736    15862 bp    DNA        BCT         03-SEP-1998
LOCUS
DEFINITION Aquifex aeolicus section 68 of 109 of the complete genome.
ACCESSION AE000736 AE000657
NID q2983763
VERSION AE000736.1 GI:2983763
KEYWORDS
SOURCE
ORGANISM Aquifex aeolicus.
REFERENCE Eubacteria; Aquificales; Aquificaceae; Aquifex.
AUTHORS Decker,T.G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L., Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Anjaj,M., Huber,R., Feldman,R.A., Short,J.M., Olson,G.J. and Swanson,R.V. The complete genome of the hyperthermophilic bacterium Aquifex aeolicus Nature 392 (6674), 353-358 (1998)
TITLE
JOURNAL Nature 392 (6674), 353-358 (1998)
MEDLINE 98196666
REFERENCE 2 (bases 1 to 15862)
AUTHORS Decker,T.G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L., Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Anjaj,M., Huber,R., Feldman,R.A., Short,J.M., Olson,G.J. and Swanson,R.V. Direct Submission
TITLE Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego, CA 92121
COMMENT Putative indicates no similarity to known proteins
FEATURES Hypothetical indicates similarity to a protein of unknown function.
SOURCE Location/Qualifiers
1..15862
/organism="Aquifex aeolicus"
/strain="VF5"
/db_xref="taxon:63363"
233..937
/gene="aq_1345"
233..937
/gene="aq_1345"
/codon_start=1
/transl_table=11
/product="putative protein"
/protein_id="AAC07332.1"
/db_xref="PID:g2983773"
/db_xref="GI:2983773"
/translation="MMGLLRRLRLRVRAAHASVKELEDPKITTEBGLRELKOLKEA
MEALAVKVSQIINFEEARNNEKERAKALVKKESLLLOQOEKISPEEAEKLAELIQ

```



gene  
CDS  
939..1469  
/gene="ag\_1347"  
/gene="ag\_1347"  
939..1469  
/gene="ag\_1347"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07333.1"  
/db\_xref="pid:g2983774"  
/db\_xref="gi:2983774"  
/translation="MGFLERIKTLFKEEKEQESLSNPLERKKKDDIIEIDGTEW  
VTDVALDVGASKEWEIRSRSGRFSLSEEGKITFFEEIDPELDPDGERRKRG  
RKEPEVTYTKGRKRLRYAKGAKYIKNLSESPYPTIWFERSEDGMIDLEIDEYEIAV  
KRELOEWEEIESILTR"

gene  
CDS  
1384..1953  
/gene="ag\_1348"  
/gene="ag\_1348"  
1384..1953  
/gene="ag\_1348"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07334.1"  
/db\_xref="pid:g2983775"  
/db\_xref="gi:2983775"  
/translation="MRFGMSTRKLKPTKGENTRNGLRYAFELVSLILLEFLSCSSHKVEYK  
GDEEFELKNYENKGIYRAILENKPODEYCKVHLKDLIKGEYILLPSYKVAREK  
ISTKYCELFSGRFLVAGERTPPSAKLYVTVELKEGLFMFLERVARAOGGGGGG  
IIFLLPDMGGGYNEBGRYRTPTORICK"  
1950..2756  
/gene="aq\_1349"  
/gene="aq\_1349"  
1950..2756  
/gene="aq\_1349"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07335.1"  
/db\_xref="pid:g2983776"  
/db\_xref="gi:2983776"  
/translation="MSGFLIALFWYIFSKYVEDVLFPPEAKTIEREIFGNKNALISY  
AGYPLGLAPRSYVFYESLPREVLYILFYVSFTLLGVYIPDLIFLKRIDLKEETLNG  
NAGAOTIGITFLSGLILISAFWRKRESFLISVYSILSYLGVMFLISTLLMSRL  
KLNFEEKAKNFENSALVGSITLGVSVLGAISGERMSGILFDSTLVLYFVSSOV  
LMVIYVVEFLFEKRVLISSEVENNLASLIASFIAAFITLAMG"  
2753..4234  
/gene="aq\_1350"  
/gene="aq\_1350"  
2753..4234  
/gene="aq\_1350"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC07337.1"  
/db\_xref="pid:g2983778"  
/db\_xref="gi:2983778"  
/translation="MSYSTALKLCIFTGFAGIVAEXSLATLATYLIGNAVLOWSVYI  
STFLSMGSHASHRYIPDDKTPLAFVAELFLSLVPFSVPIAHANNFNHLOTYI  
VLSYSGISGLTFPLAVRTNNWEEKNISSVLEBDYGSPADLIYAIFLRPL  
GLPLTAIAGGFNLISAFLLVKVLPKFKFLFLAIITFFELATYAVGHKRTIIIEEDK  
FGEGBIHIEBOTPYOKIVLTRGRAYSLYDGHQESTLDERKAHEMLVHPASPFLR  
YEKAILLOGGDRLARELRYKYPFGEIHLVDLDPMMIERSKNLMYRKINENSEFDTEL  
KVESDALENFYAKTEKEYDVLDLIDRPRESSAWYSLPEYMSLKMKLEDGFIHQ  
ADDTFRKRVCSILTGRKKAIFYAYPLVYIIPFGEMGWIGSKEDPLNENPELKER  
TEFLNRPAALAYTIKGLCECPNVEVNILLKPULIYYIKQN"  
4260..5306  
/note="ag\_1351"  
/gene="phoH"  
/gene="phoH"  
4260..5306  
/gene="phoH"  
/codon\_start=1  
/transl\_table=11  
/product="phosphate starvation-inducible protein"

gene  
/protein\_id="AAC07324.1"  
/db\_xref="pid:92983765"  
/db\_xref="gi:2983765"  
/translation="MEKYLAEERIDLGRYDERFYAIVRGDENTIKFSEKFPDYAIR  
GELLIIKEEDKVRREFVFLKAVLRLEHEHETLAEVKEVAKOFLKSEVAKTIEEQPG  
REERLESEVILLITRRKAIYVKTETOKRYVEALKNDIYVIGIPAGTGYKTYLAMA  
LOHLEGVNKTILTRPAVEAGEKIGFLPGTIAKRVHYLPYLDYALDYWDYKAVY  
MLERNIIETIAPLAFWRGRTLDNFTIIDEANANTRDOKMMLTRTIGGSAVITGDY  
QTDLKKEESGILEVEVLNKIEGIEFFEFKQDYVRHPYAKIITNAIEQERAKEOS  
SKRTLEESKRGOD"  
5230.5682  
/gene="ag\_1354"  
5230.5682  
/gene="ag\_1354"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC07338.1"  
/db\_xref="pid:92983779"  
/db\_xref="gi:2983779"  
/translation="MSSYTRQNRNRYLKLKKRYRKDKIEKMAELASALGNVELS  
VYITDDQIRIELNKYTRKKRKPDTVLSEPMGEFGYKILGDVVISODTIERQARELGS  
HSLSEEVRLYVHGIVHLIGYDHEKGGESEKKFPELENYVYLSKSKAL"  
complement(5649.6479)  
/note="ag\_1355"  
/gene="11pa"  
complement(5649.6479)  
/gene="11pa"  
/codon\_start=1  
/transl\_table=11  
/product="lipidic acid synthetase"  
/protein\_id="AAC07325.1"  
/db\_xref="pid:92983766"  
/db\_xref="gi:2983766"  
/translation="MMKPVLPFHSKLYEYKRLKRSRLTYTVEESRCPNISRCFENKTA  
TPTMIGNCTRCACFNCEKSPKGVNDEPEYRIILEAVTKLKYVITSTRDPLD  
GCASTFACIRLKENIEDIRYEVILPDERGNKALEVYERKEPVVNHANETVPRLL  
PSVRIGAYKSLNLTLMKSEKIDKSVYTKSALLIGFERKEEVKIRVMEKRLSDVCDL  
VLGYQYQPSLKHHPVVKYSESEKFEELIGYEMGFYVVSQPNARSSYAFESLST  
"

gene  
6505.6939  
/gene="ag\_1356"  
6505.6939  
/gene="ag\_1356"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07336.1"  
/db\_xref="pid:92983777"  
/db\_xref="gi:2983777"  
/translation="MGKVIYPPGEGPEPKKEKELKTEPLEEDFNNOVIEYFKDLKT  
LYPSQVKNLIVIGIGEEKYGDKEKPYVGLIMENHKKHSPLVSSFYVALKQDKLEISF  
LEAKNNYDMKIMKMTYTKDFHTELMNGLEFLKEREKEGI"  
7056.8462  
/note="ag\_1357"  
/gene="cyda"  
7056.8462  
/gene="cyda"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome oxidase d subunit I"  
/protein\_id="AAC07328.1"  
/db\_xref="pid:92983769"  
/db\_xref="gi:2983769"  
/translation="MDAVLARIQIOPALTAHFLFVPTTIGLSLLAIINTKAREED  
DERHILRISEIEFTTFLALINFAGVAGVILEEFTNMSSEKTDGDFEGAPLAE  
ALFAPELSEVFGLVLEFGDRVSREYFSTWMAVIGTNSALMILVANSWOOTPKGF  
YLEGKATLASPEAVVNYSTPYRPHMFLAGVITGAIFVAGISAVYILKRHLDPK  
ESFRILAIIGMNASLLOIAGDLGYOVAROPILKAMMEDKMETESGAGLETPGIE  
IPKLSILAHYDPDAKYIGIDIVOTQRIKAGEVEDELEPELMDPPROPAGROSLY  
ITEEBIPFTWIPFNSLSRGWVLEGFPAFTISLGIITFVODITEDRMWLLDFSTDL  
PLVANIILMWIAEYGRQWVYVYILTKRDAASPLPAQOILTSIILFSSISYIFVFLIFL  
GLMWOKVIGKRRPEEVA"

gene  
/protein\_id="AAC07324.1"  
/db\_xref="pid:92983765"  
/db\_xref="gi:2983765"  
/translation="MEKYLAEERIDLGRYDERFYAIVRGDENTIKFSEKFPDYAIR  
GELLIIKEEDKVRREFVFLKAVLRLEHEHETLAEVKEVAKOFLKSEVAKTIEEQPG  
REERLESEVILLITRRKAIYVKTETOKRYVEALKNDIYVIGIPAGTGYKTYLAMA  
LOHLEGVNKTILTRPAVEAGEKIGFLPGTIAKRVHYLPYLDYALDYWDYKAVY  
MLERNIIETIAPLAFWRGRTLDNFTIIDEANANTRDOKMMLTRTIGGSAVITGDY  
QTDLKKEESGILEVEVLNKIEGIEFFEFKQDYVRHPYAKIITNAIEQERAKEOS  
SKRTLEESKRGOD"  
5230.5682  
/gene="ag\_1354"  
5230.5682  
/gene="ag\_1354"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC07338.1"  
/db\_xref="pid:92983779"  
/db\_xref="gi:2983779"  
/translation="MSSYTRQNRNRYLKLKKRYRKDKIEKMAELASALGNVELS  
VYITDDQIRIELNKYTRKKRKPDTVLSEPMGEFGYKILGDVVISODTIERQARELGS  
HSLSEEVRLYVHGIVHLIGYDHEKGGESEKKFPELENYVYLSKSKAL"  
complement(5649.6479)  
/note="ag\_1355"  
/gene="11pa"  
complement(5649.6479)  
/gene="11pa"  
/codon\_start=1  
/transl\_table=11  
/product="lipidic acid synthetase"  
/protein\_id="AAC07325.1"  
/db\_xref="pid:92983766"  
/db\_xref="gi:2983766"  
/translation="MMKPVLPFHSKLYEYKRLKRSRLTYTVEESRCPNISRCFENKTA  
TPTMIGNCTRCACFNCEKSPKGVNDEPEYRIILEAVTKLKYVITSTRDPLD  
GCASTFACIRLKENIEDIRYEVILPDERGNKALEVYERKEPVVNHANETVPRLL  
PSVRIGAYKSLNLTLMKSEKIDKSVYTKSALLIGFERKEEVKIRVMEKRLSDVCDL  
VLGYQYQPSLKHHPVVKYSESEKFEELIGYEMGFYVVSQPNARSSYAFESLST  
"

gene  
6505.6939  
/gene="ag\_1356"  
6505.6939  
/gene="ag\_1356"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07336.1"  
/db\_xref="pid:92983777"  
/db\_xref="gi:2983777"  
/translation="MGKVIYPPGEGPEPKKEKELKTEPLEEDFNNOVIEYFKDLKT  
LYPSQVKNLIVIGIGEEKYGDKEKPYVGLIMENHKKHSPLVSSFYVALKQDKLEISF  
LEAKNNYDMKIMKMTYTKDFHTELMNGLEFLKEREKEGI"  
7056.8462  
/note="ag\_1357"  
/gene="cyda"  
7056.8462  
/gene="cyda"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome oxidase d subunit I"  
/protein\_id="AAC07328.1"  
/db\_xref="pid:92983769"  
/db\_xref="gi:2983769"  
/translation="MDAVLARIQIOPALTAHFLFVPTTIGLSLLAIINTKAREED  
DERHILRISEIEFTTFLALINFAGVAGVILEEFTNMSSEKTDGDFEGAPLAE  
ALFAPELSEVFGLVLEFGDRVSREYFSTWMAVIGTNSALMILVANSWOOTPKGF  
YLEGKATLASPEAVVNYSTPYRPHMFLAGVITGAIFVAGISAVYILKRHLDPK  
ESFRILAIIGMNASLLOIAGDLGYOVAROPILKAMMEDKMETESGAGLETPGIE  
IPKLSILAHYDPDAKYIGIDIVOTQRIKAGEVEDELEPELMDPPROPAGROSLY  
ITEEBIPFTWIPFNSLSRGWVLEGFPAFTISLGIITFVODITEDRMWLLDFSTDL  
PLVANIILMWIAEYGRQWVYVYILTKRDAASPLPAQOILTSIILFSSISYIFVFLIFL  
GLMWOKVIGKRRPEEVA"















[illegible]







- |                      |  |
|----------------------|--|
| JOURNAL<br>AUTHORS   | 9 (sites)<br>Katinka,M., Cossart,P., Sibilli,L., Saint-Gilrons,I.,<br>Chalvagnac,M.A., Le Bras,G., Cohen,G.N. and Yaniv,M.   |
| TITLE                | Nucleotide sequence of the thrA gene of Escherichia coli   |
| JOURNAL<br>REFERENCE | Proc. Natl. Acad. Sci. U.S.A. 77 (10), 5730-5733 (1980)  |
| MEDLINE              | 81077247   |
| JOURNAL<br>AUTHORS   | 10 (sites)<br>Mackie,G.A.  |
| TITLE                | Nucleotide sequence of the gene for ribosomal protein S20 and its flanking regions   |
| JOURNAL<br>REFERENCE | J. Biol. Chem. 256 (15), 8177-8182 (1981)  |
| MEDLINE              | 81264207   |
| JOURNAL<br>AUTHORS   | 11 (sites)<br>Cossart,P., Katinka,M. and Yaniv,M.  |
| TITLE                | Nucleotide sequence of the thrB gene of E. coli, and its two adjacent regions; the thrAB and thrBC junctions   |
| JOURNAL<br>REFERENCE | Nucleic Acids Res. 9 (2), 339-347 (1981)   |
| MEDLINE              | 81150470   |
| JOURNAL<br>AUTHORS   | 12 (sites)<br>Lee,N.L., Gielow,W.O. and Wallace,R.G.   |
| TITLE                | Mechanism of arae autoregulation and the domains of two overlapping promoters, pc and PBAD, in the L-arabinose regulatory region of Escherichia coli                             |
| JOURNAL<br>REFERENCE | Proc. Natl. Acad. Sci. U.S.A. 78 (2), 752-756 (1981)   |
| MEDLINE              | 81199399   |
| JOURNAL<br>AUTHORS   | 13 (sites)<br>Stoner,C.M. and Schleif,R.   |
| TITLE                | Is the amino acid but not the nucleotide sequence of the Escherichia coli araC gene conserved?   |
| JOURNAL<br>REFERENCE | J. Mol. Biol. 154 (4), 649-652 (1982)  |
| MEDLINE              | 82216830   |
| JOURNAL<br>AUTHORS   | 14 (sites)<br>Gilson,E., Nikaido,H. and Hofnung,M.   |
| TITLE                | Sequence of the malX gene in E.coli K12  |
| JOURNAL<br>REFERENCE | Nucleic Acids Res. 10 (22), 7449-7458 (1982)   |
| MEDLINE              | 83116968   |
| JOURNAL<br>AUTHORS   | 15 (sites)<br>Parrot,C., Cossart,P., Saint-Gilrons,I. and Cohen,G.N.   |
| TITLE                | Nucleotide sequence of thrC and of the transcription termination region of the threonine operon in Escherichia coli K12  |
| JOURNAL<br>REFERENCE | Nucleic Acids Res. 11 (21), 7331-7345 (1983)   |
| MEDLINE              | 84069770   |
| JOURNAL<br>AUTHORS   | 16 (sites)<br>Bouvier,J., Richard,C., Richard,F., Patte,J.C. and Stragier,P.   |
| TITLE                | Nucleotide sequence and expression of the Escherichia coli daps gene   |
| JOURNAL<br>REFERENCE | J. Biol. Chem. 259 (23), 14829-14834 (1984)  |
| MEDLINE              | 85054974   |
| JOURNAL<br>AUTHORS   | 17 (sites)<br>Bardwell,J.C. and Craig,E.A.   |
| TITLE                | Major heat shock gene of Drosophila and the Escherichia coli heat-inducible dnaK gene are homologous   |
| JOURNAL<br>REFERENCE | Proc. Natl. Acad. Sci. U.S.A. 81 (3), 848-852 (1984)   |
| MEDLINE              | 84144800   |
| JOURNAL<br>AUTHORS   | 18 (sites)<br>Imnis,M.A., Tokunaga,M., Williams,M.E., Loranger,J.M., Chang,S.Y., Chang,S. and Wu,H.C.  |
| TITLE                | Nucleotide sequence of the Escherichia coli prolipoprotein signal peptidase (isp) gene   |
| JOURNAL<br>REFERENCE | Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3708-3712 (1984)  |
| MEDLINE              | 84222028   |
| JOURNAL<br>AUTHORS   | 19 (sites)<br>Bouvier,J., Patte,J.C. and Stragier,P.   |
| TITLE                | Multiple regulatory signals in the control region of the Escherichia coli carAB operon   |
| JOURNAL<br>REFERENCE | Proc. Natl. Acad. Sci. U.S.A. 81 (13), 4139-4143 (1984)  |
| MEDLINE              | 84248073   |
| JOURNAL<br>AUTHORS   | 20 (sites)<br>Chong,P., Hui,I., Loo,T. and Gillam,S.   |
| TITLE                | Structural analysis of a new GC-specific insertion element IS186   |
| JOURNAL<br>REFERENCE | FEMS Lett. 192 (1), 47-52 (1985)   |
| MEDLINE              | 86030702   |
| JOURNAL<br>REFERENCE | 21 (sites)   |
| JOURNAL<br>AUTHORS   | Kamo,Y., Iin,C.K., Regue,M. and Wu,H.C.  |
| TITLE                | Characterization of the iles-lsp operon in Escherichia coli. Identification of an open reading frame upstream of the iles gene and potential promoter(s) for the iles-lsp operon |
| JOURNAL<br>REFERENCE | J. Biol. Chem. 260 (9), 5616-5620 (1985)   |
| MEDLINE              | 85182715   |
| JOURNAL<br>AUTHORS   | 22 (sites)<br>Friedberg,D., Rosenthal,E.R., Jones,J.W. and Calvo,J.M.  |
| TITLE                | Characterization of the 3' end of the leucine operon of Salmonella typhimurium   |
| JOURNAL<br>REFERENCE | Mol. Gen. Genet. 199 (3), 486-494 (1985)   |
| MEDLINE              | 85295470   |
| JOURNAL<br>AUTHORS   | 23 (sites)<br>Cowling,D.W., Bardwell,J.C., Craig,E.A., Woolford,C., Hendrix,R.W. and Gross,C.A.  |
| TITLE                | Consensus sequence for Escherichia coli heat shock gene promoters  |
| JOURNAL<br>REFERENCE | Proc. Natl. Acad. Sci. U.S.A. 82 (9), 2679-2683 (1985)   |
| MEDLINE              | 85190560   |
| JOURNAL<br>AUTHORS   | 24 (sites)<br>Sekiguchi,T., Ortega-Cesena,J., Nosch,Y., Ohashi,S., Tsuda,K. and Kanaya,S.  |
| TITLE                | DNA and amino-acid sequences of 3-isopropylmalate dehydrogenase of Bacillus coagulans. Comparison with the enzymes of Saccharomyces cerevisiae and Thermus thermophilus          |
| JOURNAL<br>REFERENCE | Biochim. Biophys. Acta 867, 36-44 (1986)   |
| MEDLINE              | 25 (sites)   |
| JOURNAL<br>AUTHORS   | Lee,N., Gielow,W., Martin,R., Hamilton,E. and Fowler,A.  |
| TITLE                | The organization of the arabAD operon of Escherichia coli  |
| JOURNAL<br>REFERENCE | Gene 47 (2-3), 231-244 (1986)  |
| MEDLINE              | 87163495   |
| JOURNAL<br>AUTHORS   | 26 (sites)<br>Ohki,M., Tamura,F., Nishimura,S. and Uchida,H.   |
| TITLE                | Nucleotide sequence of the Escherichia coli dnaJ gene and purification of the gene product   |
| JOURNAL<br>REFERENCE | J. Biol. Chem. 261 (4), 1778-1781 (1986)   |
| MEDLINE              | 86111849   |
| JOURNAL<br>AUTHORS   | 27 (sites)<br>Blanchin-Roland,S., Blanquet,S., Schmitter,J.M. and Fayat,G.   |
| TITLE                | The gene for Escherichia coli diadenosine tetraphosphatase is located immediately clockwise to folA and forms an operon with ksgA  |
| JOURNAL<br>REFERENCE | Mol. Gen. Genet. 205 (3), 515-522 (1986)   |
| MEDLINE              | 87172305   |
| JOURNAL<br>AUTHORS   | 28 (sites)<br>Mackie,G.A.  |
| TITLE                | Structure of the DNA distal to the gene for ribosomal protein S20 in Escherichia coli K12: presence of a strong terminator and an ISI element                                    |
| JOURNAL<br>REFERENCE | Nucleic Acids Res. 14 (17), 6965-6981 (1986)   |
| MEDLINE              | 87016337   |
| JOURNAL<br>AUTHORS   | 29 (sites)<br>Birnbauum,M.J., Haspel,H.C. and Rosen,O.M.   |
| TITLE                | Cloning and characterization of a cDNA encoding the rat brain glucose-transporter protein  |
| JOURNAL<br>REFERENCE | Proc. Natl. Acad. Sci. U.S.A. 83 (16), 5784-5788 (1986)  |
| MEDLINE              | 86287338   |
| JOURNAL<br>AUTHORS   | 30 (sites)<br>Nonet,M.L., Marvel,C.C. and Tolian,D.R.  |
| TITLE                | The hist-purF region of the Escherichia coli K-12 chromosome. Identification of additional genes of the hist and purF operons  |
| JOURNAL<br>REFERENCE | J. Biol. Chem. 262 (25), 12209-12217 (1987)  |
| MEDLINE              | 87308226   |
| JOURNAL<br>AUTHORS   | 31 (sites)<br>Gronger,P., Manian,S.S., Reilander,H., O'Connell,M., Priefer,U.B. and Puhler,A.  |
| TITLE                | Organization and partial sequence of a DNA region of the Rhizobium leguminosarum symbiotic plasmid pRL601 containing the genes fixABC, nifH, nifB and a novel open reading frame |
| JOURNAL<br>REFERENCE | Nucleic Acids Res. 15 (1), 31-49 (1987)  |
| MEDLINE              | 87146339   |
| Query Match          | 10.6%; Score 283; DB 17; Length 111401;  |
| Note:                | remainder of annotations omitted.  |















(a) (b) (c) (d) (e) (f)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

MPsrch_pp  protein - protein database search, using Smith-Waterman algorithm

Run on:      Wed Nov 24 00:53:13 1999;      MasPar time 0.42 Seconds
Tabular output not generated.      196.757 Million cell updates/sec

```

```
Title: >US-09-103-287-4
Description: (1-215) from US09103287.pep
Perfect Score: 1495
Sequence: 1 FKSSDDIYAQIFQITDKGTA.....GDIQKLQNAVLDRIGKKNAP 215
```

Scoring table: PAM 150  
Gap 11

Searched: 1479 seqs, 387401 residues

```
Post-processing: Minimum Match 08
Listing first 45 summaries
```

Database: HIV-AA8  
1:ALL

Statistics: Mean 42.724; Variance 87.119; scale 0.490

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	78	5.2	489	1	GAGSVI57	HIVV57, GAG POLYPROT	1.296+00
2	73	4.9	385	1	GAGSVI32	HIVV32, GAG POLYPROT	4.266+00
3	73	4.9	486	1	GAGSVI415	HIVV415, GAG POLYPRO	4.266+00
4	73	4.9	490	1	GAGSVI415	HIVV415, GAG POLYPRO	4.266+00
5	72	4.8	853	1	ENVELOPE\$A	HIVV423, GAG POLYPRO	4.266+00
6	70	4.7	242	1	GAGSVIC144	HIVV423, GAG POLYPRO	5.366+00
7	71	4.7	264	1	GAGSVI354	HIVV423, GAG POLYPRO	8.396+00
8	70	4.7	304	1	GAGSVI359	HIVV423, GAG POLYPRO	6.726+00
9	71	4.7	489	1	GAGSVI245	HIVV245, GAG POLYPRO	8.396+00
10	70	4.7	492	1	GAGSVI245	HIVV245, GAG POLYPRO	8.396+00
11	70	4.7	1033	1	POLSMVP518	HIVV518, POL POLYPRO	1.056+01
12	69	4.6	486	1	GAGSDI258	HIVV258, GAG POLYPRO	1.056+01
13	69	4.6	488	1	GAGSVI310	HIVV310, GAG POLYPRO	1.056+01
14	69	4.6	1045	1	POLSVLCVCG	HIVV310, GAG POLYPRO	1.056+01
15	68	4.5	91	1	GAGSM251	HIVV251, GAG	1.306+01
16	68	4.5	103	1	GAGSF10S1	HIVV10S1, GAG	1.306+01
17	67	4.5	486	1	GAGSVI4231	HIVV4231, GAG POLYPRO	1.606+01
18	68	4.5	487	1	GAGSVI174	HIVV174, GAG POLYPRO	1.306+01
19	68	4.5	489	1	GAGSVI112	HIVV112, GAG POLYPROT	1.306+01
20	68	4.5	489	1	GAGSVI112	HIVV112, GAG POLYPROT	1.306+01
21	67	4.5	489	1	GAGSVI112	HIVV112, GAG POLYPROT	1.306+01
22	67	4.5	489	1	GAGSVI112	HIVV112, GAG POLYPROT	1.306+01
23	67	4.5	489	1	GAGSVI112	HIVV112, GAG POLYPROT	1.306+01

24	67	4.5	497	1	GAGSTB1321	HVITB1321, GAG	1.60e+01
25	67	4.5	886	1	ENVELOPEPSS	SIVSM9, ENVELOPE POL	1.60e+01
26	67	4.5	1027	1	POLSCPZ	SIVCPZ, POL POLYPROTE	1.60e+01
27	66	4.4	486	1	GAGSGD41	HVIG141, GAG	1.97e+01
28	66	4.4	1009	1	POLSMDCB1	SIVMDCB1, POL	1.97e+01
29	65	4.3	93	1	ENVELOPE_U	HVUD04508, ENVELOPE	2.42e+01
30	64	4.3	211	1	GAGSB1200	HVIBZ200, GAG POLYPRO	2.96e+01
31	64	4.3	481	1	GAGSB120	HVIC120, GAG	2.96e+01
32	64	4.3	486	1	GAGSV169	HVIV169, GAG POLYPROT	2.96e+01
33	64	4.3	488	1	GAGSV0268	HVIV268, GAG POLYPRO	2.96e+01
34	64	4.3	500	1	GAGSBCSG3C	HVIVCSG3C, GAG POLYPR	2.96e+01
35	64	4.3	500	1	GAGSNL43	HVNL43, GAG POLYPROT	2.96e+01
36	65	4.3	505	1	GAGSHSPGAG	HVIAL, GAG POLYPROTE	2.96e+01
37	65	4.3	811	1	GAGSHSPGAG	HSPGAGPOL, GAG	2.96e+01
38	65	4.3	852	1	ENVELOPE_U	HVIV08443, ENVELOPE	2.42e+01
39	65	4.3	1003	1	POLSYU2	HVIV2, POL POLYPROTE	2.96e+01
40	65	4.3	1036	1	POLSSKX	SIVSKX, POL PROTEIN	2.42e+01
41	63	4.2	484	1	GAGSC41	HVIV14, GAG POLYPROTE	3.61e+01
42	63	4.2	491	1	GAGSPH136	HVPH136, GAG POLYPRO	3.61e+01
43	63	4.2	1002	1	POLSEPL1	HVIV1, POL POLYPROTE	3.61e+01
44	63	4.2	1003	1	POLSDJ1	HVIVD1, POL POLYPROTE	3.61e+01
45	63	4.2	1015	1	POLSTLAI	HVIVAI, POL POLYPROTE	3.61e+01

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	489	AA
ID	GAGSV157				
AC	L11794				
DT	20-APR-1993				
DE	HIV157, GAG POLYPROTEIN				
FT	PEPT				
CC	-1-5':Y 3':N				
CC	TRANSLATED USING PHASE 1				
CC	SEQUENCE 489				
CC	AA: 54331				
CC	MM:				

```
Query Match      5.2%; Score 78; DB 1; Length 489;
Best Local Similarity 22.0%; Pred. No. 1.29e+00;
```

```
Db 75 lkslnftvaivycvqhrievktdkdaIdkieeiqlkksqxt 115
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 44 LNL-AVAISY-L-EKLDVTNIKEALETFEGCVKRRENETT 81
```

SEQUENCE	AA	MM	MM
TRANSLATED USING PHASE 1	AA: 42556	MM:	
PEPT	1	>385	
-1-5':Y	3':N		
GAG POLYPROTEIN (PREMATURE TERMINATION)			

Query Match	4.98;	Score 73;	DB 1;	Length 385;
Best Local Similarity	27.08;	Pred. No. 4.26e+00;		
Matches	10;	Conservative	15;	Mismatches 11;
				Indels 1;
				Gaps 1

```
D0      80 ntavvl-fcvhgridvkdtkaldkiseelknksqka 115  
       ||::||:::|::||::||::||::||::||::||:::  
QY      45 NALAVIAISTYLEKLDVTNIKEALETFEGVKKRFNETT 81
```

RESULT 3

ID	GAGSVI415		PRELIMINARY;	PRT;	486	AA
AC	L11791					
DT	20-APR-1993					
DE	HIVVI415,	GAG POLYPEPTIDEIN				
FE	PEPT	1	>486			GAG POLYPEPTIDEIN
CC	-1-5::Y	3::N				
CC	TRANSLATED USING PHASE 1					
CQ	SEQUENCE	486	AA;	53947	NN;	



Query Match 4.9%; Score 73; DB 1; Length 486;  
Best Local Similarity 30.8%; Pred. No. 4.26e+00;  
Matches 8; Conservative 10; Mismatches 8; Indels 0; Gaps 0;

Db 90 grldvkdtklealdkleeqknkqkt 115  
56 EKLDVTNIKEALTEFGVKRRNETT 81

RESULT 4  
ID GAGSLBV23 PRELIMINARY; PRT; 490 AA

AC L11777  
DT 20-APR-1993  
DE HIVLBV23, GAG POLYPROTEIN  
FT PEPT 1 >490 GAG POLYPROTEIN  
CC -1-5',Y 3',N  
CC TRANSLATED USING PHASE 1  
SQ SEQUENCE 490 AA; 54634 MW;

Query Match 4.9%; Score 73; DB 1; Length 490;  
Best Local Similarity 27.8%; Pred. No. 4.26e+00;  
Matches 10; Conservative 14; Mismatches 9; Indels 3; Gaps 3;

Db 75 lkslnvavlycvhqrldvkdtklealdtkieeiqkk 110  
44 LNAL-AVAIAISY-L-EKLDVTNIKEALTEFGVKRR 76

RESULT 5  
ID ENVELOPESADA PRELIMINARY; PRT; 853 AA

AC M60472  
DT 16-FEB-1989  
DE HIVADA, ENVELOPE POLYPROTEIN  
FT PEPT 1 853 ENVELOPE POLYPROTEIN  
CC -1-5',Y 3',Y  
CC TRANSLATED USING PHASE 1  
SQ SEQUENCE 853 AA; 96998 MW;

Query Match 4.8%; Score 72; DB 1; Length 853;  
Best Local Similarity 30.9%; Pred. No. 5.36e+00;  
Matches 21; Conservative 12; Mismatches 33; Indels 3; Gaps 3;

Db 121 ltplyv-clnctdlrvnnlnnssgmrgelknscfnlttslrckvkkdyalfyrlavp 179  
44 LNALAVIAISYLEKLDVTNIKEALTEF-GVKR-RNETTIANQVIYDDYAHHPREISAT 101

Db 180 idndntsy 187  
102 IDTARKKY 109

RESULT 6  
ID GAGSIC144 PRELIMINARY; PRT; 242 AA

AC L11767  
DT 20-APR-1993  
DE HIVIC144, GAG  
FT PEPT 1 >242 GAG (PREMATURE TERMINATION)  
CC -1-5',Y 3',N  
CC TRANSLATED USING PHASE 1  
SQ SEQUENCE 242 AA; 26857 MW;

Query Match 4.7%; Score 70; DB 1; Length 242;  
Best Local Similarity 19.0%; Pred. No. 8.39e+00;  
Matches 11; Conservative 18; Mismatches 27; Indels 2; Gaps 2;

Db 71 gseevkslnfatlyclhqrldvkdtklealdkleeiqnkskqkqgaavtsssqn 128  
39 GDHVTNLNALAVIAISY-L-EKLDVTNIKEALTEFGVKRRNETTIANQVIYDDYAHH 94

RESULT 7  
ID GAGSVI354 PRELIMINARY; PRT; 264 AA

AC L11790  
DT 20-APR-1993  
DE HIVI354, GAG POLYPROTEIN  
FT PEPT 1 >264 GAG POLYPROTEIN (PREMATURE TERMINATION)  
CC -1-5',Y 3',N  
CC TRANSLATED USING PHASE 1  
SQ SEQUENCE 264 AA; 29206 MW;

Query Match 4.7%; Score 71; DB 1; Length 264;  
Best Local Similarity 22.0%; Pred. No. 6.72e+00;  
Matches 11; Conservative 16; Mismatches 20; Indels 3; Gaps 3;

Db 75 lkslnvavlycvhqrldvkdtklealdkleeiqnkskqkqgaataen 124  
44 LNAL-AVAIAISY-L-EKLDVTNIKEALTEFGVKRRNETTIANQVIYDD 90

RESULT 8  
ID GAGSCI59 PRELIMINARY; PRT; 304 AA

AC L11759  
DT 20-APR-1993  
DE HIVCI59, GAG POLYPROTEIN  
FT PEPT 1 >304 GAG POLYPROTEIN (PREMATURE TERMINATION)  
CC -1-5',Y 3',N  
CC TRANSLATED USING PHASE 1  
SQ SEQUENCE 304 AA; 34039 MW;

Query Match 4.7%; Score 70; DB 1; Length 304;  
Best Local Similarity 27.5%; Pred. No. 8.39e+00;  
Matches 11; Conservative 15; Mismatches 13; Indels 1; Gaps 1;

Db 77 slfnalavlycvh-grldvkdtklealdkleeqknksqkt 115  
42 TVNLALAVIAISYLEKLDVTNIKEALTEFGVKRRNETT 81

RESULT 9  
ID GAGSTN245 PRELIMINARY; PRT; 489 AA

AC L11762  
DT 27-MAY-1993  
DE HIVTN245, GAG POLYPROTEIN  
FT PEPT 1 >489 GAG POLYPROTEIN  
CC -1-5',Y 3',N  
CC TRANSLATED USING PHASE 1  
SQ SEQUENCE 489 AA; 54658 MW;

Query Match 4.7%; Score 71; DB 1; Length 489;  
Best Local Similarity 22.9%; Pred. No. 6.72e+00;  
Matches 8; Conservative 14; Mismatches 12; Indels 1; Gaps 1;

Db 81 tvvlwcvhqrldvkdtklealdkleeqknksqkt 115  
48 AVIAISYL-EKLDVTNIKEALTEFGVKRRNETT 81

RESULT 10  
ID GAGSKI24 PRELIMINARY; PRT; 492 AA

AC L11769  
DT 20-APR-1993  
DE HIVKI24, GAG POLYPROTEIN  
FT PEPT 1 >492 GAG POLYPROTEIN  
CC -1-5',Y 3',N  
CC TRANSLATED USING PHASE 1  
SQ SEQUENCE 492 AA; 54739 MW;

Query Match 4.7%; Score 70; DB 1; Length 492;  
Best Local Similarity 24.4%; Pred. No. 8.39e+00;  
Matches 10; Conservative 17; Mismatches 13; Indels 1; Gaps 1;

Db 76 tslnavavlycvh-grldvkdtklealdkleeqknksqkt 115  
41 HTVNLALAVIAISYLEKLDVTNIKEALTEFGVKRRNETT 81



```
RESULT 11
ID POLSMVP5180 PRELIMINARY; PRT; 1033 AA
AC L20571
DT 23-JUN-1993
DE HIVVP5180, POL POLYPROTEIN
FT PEPT <1 1033
CC -1-5':N 3':Y
CC TRANSLATED USING PHASE 1
SQ SEQUENCE 1033 AA; 116942 MW;

Query Match
Best Local Similarity 20.4%; Score 70; DB 1; Length 1033;
Pred. No. 8.39e+00;
Matches 11; Conservative 19; Mismatches 23; Indels 1; Gaps 1;

Db 677 vlgissqptgsdspiqllelkkeryllw-pahkgysgnekidklysk 729
QY 114 VVAVFQPHFRTQAFLEFASLSKADRVFCEIGSIRENTGALTLODLIDK 167

RESULT 12
ID GAGSDJ258 PRELIMINARY; PRT; 486 AA
AC L11763
DT 27-MAY-1993
DE HIVDJ258, GAG POLYPROTEIN
FT PEPT 1 >486
CC -1-5':Y 3':N
CC TRANSLATED USING PHASE 1
SQ SEQUENCE 486 AA; 54093 MW;

Query Match
Best Local Similarity 26.9%; Score 69; DB 1; Length 486;
Pred. No. 1.05e+01;
Matches 7; Conservative 11; Mismatches 8; Indels 0; Gaps 0;

Db 90 gridktdkealdkieavqnskqkt 115
QY 56 EKLDVTNIKEALETFGCVKRRFNETT 81

RESULT 13
ID GAGSVI310 PRELIMINARY; PRT; 488 AA
AC L11786
DT 20-APR-1993
DE HIVVI310, GAG POLYPROTEIN
FT PEPT 1 >488
CC -1-5':Y 3':N
CC TRANSLATED USING PHASE 1
SQ SEQUENCE 488 AA; 54661 MW;

Query Match
Best Local Similarity 27.8%; Score 69; DB 1; Length 488;
Pred. No. 1.05e+01;
Matches 10; Conservative 12; Mismatches 11; Indels 3; Gaps 3;

Db 75 lrslnfclavlycvbqrlvdktdkealdkieiqnk 110
QY 44 LNAL-AVIAISY-L-EKLDVTNIKEALETFGCVKRR 76

RESULT 14
ID POLSVLVCG PRELIMINARY; PRT; 1045 AA
AC M10608; M18039;
DT 01-MAR-1989
DE VLNGG, POL POLYPROTEIN PRECURSOR
FT PEPT 5008 1 1045 POL
CC -1-5':N 3':Y
CC TRANSLATED USING PHASE 1
SQ SEQUENCE 1045 AA; 119999 MW;

Query Match
Best Local Similarity 26.7%; Score 69; DB 1; Length 1045;
Pred. No. 1.05e+01;
Matches 16; Conservative 19; Mismatches 22; Indels 3; Gaps 3;

Db 33 lsegekweg-vhlykdkmklgtivlatpsvrylgdmme-1g-1gllmanleekkip 89
```

```
QY 25 VDGEFYDHLSPQYGDHTVLMNALAVIAISYLEKLDVTNIKEALETFGCVKRRFNETTIAN 84

RESULT 15
ID GAGSM251 PRELIMINARY; PRT; 91 AA
AC L21492
DT 09-SEP-1993
DE HIVM251, GAG
FT PEPT <1 >91
CC -1-5':N 3':N
CC TRANSLATED USING PHASE 1
SQ SEQUENCE 91 AA; 10138 MW;

Query Match
Best Local Similarity 43.3%; Score 68; DB 1; Length 91;
Pred. No. 1.30e+01;
Matches 13; Conservative 7; Mismatches 7; Indels 3; Gaps 3;

Db 12 lrslyncfclavlycvbqrlvdktdkealeki 41
QY 44 LNAL-AVIAISY-L-EKLDVTNIKEALETF 70
```

Search completed: Wed Nov 24 00:53:16 1999  
Job time : 3 secs.



**THIS PAGE BLANK (USPTO)**







ORIGIN gag cds start

Query Match 5.3%; Score 141; DB 3; Length 1480;  
Best Local Similarity 43.1%; Pred. No. 7.18e-03;  
Matches 31; Conservative 14; Mismatches 27; Indels 0; Gaps 0;

Db 269 aaagatagatgttaagagaccacaaagagcttagaagaagtagaggaagcaaacaa 328  
167 ARARATYNGAGTNCANAAATHAARGARCGNTYNGARACNTTGGNGGNTNARMGNM 226

Db 329 aaagtaagaaga 340

Qy 227 GNTTYAAYGARA 238

RESULT 2  
LOCUS HIYM251 303 bp ss-RNA VRL 09-SEP-1993  
DEFINITION Human immunodeficiency virus type 1, clone 2, ADC sample M, partial  
gag cds, p17 region.  
ACCESSION L21492  
SEGMENT 1 of 4  
SOURCE Human immunodeficiency virus type 1 (HIV-1), clone 2, ADC sample M.  
ORGANISM  
REFERENCE 1 (bases 1 to 303)  
AUTHORS Zhu,T., Mo,H., Wang,N., Nam,D.S., Cao,Y., Koup,R.A. and Ho,D.D.  
TITLE Genotypic and phenotypic characterization of HIV-1 in patients with primary infection  
JOURNAL Science 261 1179-1181 (1993)  
STANDARD full staff review  
COMMENT This sample has been taken from patient M, one of the seven patients (A,M,L,F,C,V,R), in the study(1). All patients had acute, self-limited symptomatic illness with measurable viremia followed by seroconversion. Patient M was a chronically infected male who subsequently transmitted the virus to patient F. The authors report that the sequences in this study are found to be uniformly macrophage-tropic and non-syncytium-inducing. The sequence for sample M clone 2, along with other patient M clones 1, 3-9, 11, and 12, is shown in alignment following the printed text entry. These gag sequences cluster with HIV-1 B subtype sequences. See also L21224-L21591 and L24161, L24162.

FEATURES  
CDS  
Location/Qualifiers  
<1..>303  
/product="myristylated gag protein p17"  
/gene="gag"  
/codon\_start=-1  
/organism="Human immunodeficiency virus type 1"  
/sequenced\_mol="RNA"

BASE COUNT 125 a 61 c 62 g 55 t  
ORIGIN 190 bp downstream from beginning of gag cds

Query Match 4.9%; Score 131; DB 3; Length 303;  
Best Local Similarity 41.7%; Pred. No. 7.83e-02;  
Matches 30; Conservative 14; Mismatches 28; Indels 0; Gaps 0;

Db 80 aaagatagatgttaagagaccacaaagagcttagaagaagtagaggaagcaaacaa 139  
167 ARARATYNGAGTNCANAAATHAARGARCGNTYNGARACNTTGGNGGNTNARMGNM 226

Db 140 aaagtaagaaga 151

Qy 227 GNTTYAAYGARA 238

RESULT 3  
LOCUS HIV10S1 309 bp ss-RNA VRL 09-SEP-1993  
DEFINITION Human immunodeficiency virus type 1, clone 10, ADC sample F, partial  
gag cds, p17 region.  
ACCESSION L21375  
SEGMENT 1 of 4  
SOURCE Human immunodeficiency virus type 1 (HIV-1), clone 10, ADC sample F.

ORGANISM Human immunodeficiency virus type 1  
Virus: ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.

REFERENCE 1 (bases 1 to 309)  
AUTHORS Zhu,T., Mo,H., Wang,N., Nam,D.S., Cao,Y., Koup,R.A. and Ho,D.D.  
TITLE Genotypic and phenotypic characterization of HIV-1 in patients with primary infection  
JOURNAL Science 261 1179-1181 (1993)  
STANDARD full staff review  
COMMENT This sample has been taken from patient F, one of the seven patients (A,M,L,F,C,V,R), in the study(1). All patients had acute, self-limited symptomatic illness with measurable viremia followed by seroconversion. Patient F was infected by patient M via sexual transmission. The authors report that the sequences in this study are found to be uniformly macrophage-tropic and non-syncytium-inducing. The sequence for sample F clone 10, along with other patient F clones 1-9, 11, and 12, is shown in alignment following the printed text entry. These gag sequences cluster with HIV-1 B subtype sequences. See also L21224-L21591 and L24161, L24162.

FEATURES  
CDS  
Location/Qualifiers  
<1..>309  
/product="myristylated gag protein p17"  
/gene="gag"  
/codon\_start=-1  
/translation="LQPSIQTGSSELRSLYNTAVLYCVHQKIDVKTEALEKIEE  
ONSKKKKQAOATAADTANSQVSNYPVONIGOMVHPISPTLNAMVYVEEKA  
F"

BASE COUNT 128 a 64 c 62 g 55 t  
ORIGIN 190 bp downstream from beginning of gag cds

Query Match 4.9%; Score 131; DB 3; Length 309;  
Best Local Similarity 41.7%; Pred. No. 7.83e-02;  
Matches 30; Conservative 14; Mismatches 28; Indels 0; Gaps 0;

Db 80 aaagatagatgttaagagaccacaaagagcttagaagaagtagaggaagcaaacaa 139  
167 ARARATYNGAGTNCANAAATHAARGARCGNTYNGARACNTTGGNGGNTNARMGNM 226

Db 140 aaagtaagaaga 151

Qy 227 GNTTYAAYGARA 238

RESULT 4  
LOCUS HIVPH136 1474 bp ss-RNA VRL 20-APR-1993  
DEFINITION Human immunodeficiency virus type 1, Philippines isolate PH136, gag region.  
ACCESSION L11780  
SOURCE Human immunodeficiency virus type 1 (HIV-1), Philippines isolate PH136.  
ORGANISM  
REFERENCE 1 (bases 1 to 1474)  
AUTHORS Louwagie,J.J., McCutchan,F., Brennan,T., Peeters,M., Brennan,T., Sanders-Buell,E., Eddy,G., van der Groen,G., Franssen,K., Gershy-Damet,M., Delys,R. and Burke,D.  
TITLE Phylogenetic analysis of gag genes from seventy international HIV-1 isolates provides evidence for multiple genotypes  
JOURNAL AIDS 7, 769-780 (1993)  
STANDARD  
COMMENT Kindly provided prior to publication by Henry M. Jackson Foundation Research Laboratory, 1500 East Gude Drive, Rockville, MD 20850. The PH136 gag sequence clusters with HIV-1 B subtype sequences.  
FEATURES  
CDS  
Location/Qualifiers  
1..1474  
/product="gag protein"  
/gene="gag"  
/codon\_start=-1  
/translation="MGARASYLVSGELDRWERIRLRPGKKKKYKLIKHIWASRELERF"



	source	1..1474 /organism="Human immunodeficiency virus type 1" /isolate="PH136" /cell_type="lymphocyte" /proviral /sequenced_mol="DNA" /tissue_type="blood"					
BASE COUNT	gag cds start	551 a	278 c	365 g	280 t		
ORIGIN							
Query Match		4.9%	Score 131;	DB 3;	Length 1474;		
Best Local Similarity		41.7%;	Pred. No. 7.83e-02;				
Matches	30; Conservative	14;	Mismatches 28;	Indels	0;	Gaps 0;	
Db	269	aaagaatagatgcuaaagaaccacgaagaacttagaagaatagaggagaacaaca	328				
Oy	167	ARAAATYTGATGCTNCAACNAATATHAARGAGCNYTGAACACTTTYGNGNGNTNAARMGM	226				
Db	329	aaagtaagaanaa	340				
Oy	227	GNTTYAAYGAR	238				
RESULT	5	HIVCAM1	9591 bp ss-RNA	VRL	09-JUL-1991		
LOCUS		Human immunodeficiency virus type 1,	isolate CAM1 from Cambridge,				
DEFINITION		England, complete genome.					
ACCESSION		D10112.D00917					
SOURCE		Human immunodeficiency virus type 1 (HIV-1), isolate CAM1 from Cambridge, England.					
ORGANISM		Human immunodeficiency virus type 1					
REFERENCE		McIntosh,A. and Karpas,A.					
AUTHORS		Nucleotide sequence of a Cambridge Isolate of human					
TITLE		immunodeficiency virus type 1					
JOURNAL		Thesis (1991) Cambridge University, U.K.					
STANDARD		full staff entry					
COMMENT		kindly provided in computer-readable form by Dr. Alison McIntosh, Sir William Dunn School of Pathology, University of Oxford, South Parks Rd., Oxford OX1 3RE, United Kingdom. CAM-1 sequences cluster with B subtype sequences.					
FEATURES		location/Qualifiers					
CDS		join(5832..6046,8380..8470)					
		/product="tat protein"					
		/codon_start=1					
		/translation="MEPYDPRLEPMKHGSGSPRTACNYCKQCCFHCQVCFITKGLG					
		ISYGRKKRRRRTTPPSGKTROASLSRKPSAFOSFGDPGTGPRESKKVEGETHPGD"					
		join(5971..6046,8380..8654)					
		/product="rev protein"					
		/codon_start=1					
		/translation="WAGRGSDDELLKAVALLIKLYQSNLEPPSSKTRQARRNNRRR					
		LWESQRIRAISNMILSTHLGRPAEPVPVLGIPLERLTLDSCDKGCGTGGVGSPDI					
		KRESPAVLESQTKE"					
		2..455					
		/note="5'LTR U3 region"					
		351..360					
		/note="NF-KB related site"					
		365..374					
		/note="NF-KB related site"					
		390..399					
		GC_signal					
		401..410					
		GC_signal					
		428..433					
		TATA_signal					
		456..553					
		LTR					

[illegible]



[illegible]

RESULT	7
LOCUS	HIVLBV2310 1459 bp ss-RNA
DEFINITION	Human immunodeficiency virus type 1, isolate LBV23-10 taken from a Gabonese national residing in Libreville, gag region.
ACCESSION	U11779
SOURCE	Human immunodeficiency virus type 1 (HIV-1), Libreville, Gabon isolate LBV23-10.
ORGANISM	Human Immunodeficiency virus type 1 Virus; ss-RNA enveloped viruses; Positive strand RNA virus; Viridae;

	AUTHORS	Louwaere,J.J., McCutchan,F., Brennan,T., Peeters,M., Brennan,T., Senders-Buell,E., Eddy,G., van der Groen,G., Franssen,K., Gershby-Damel,M., Deleys,R. and Burke,D.
	TITLE	Phylogenetic analysis of gag genes from seventy international HIV-1 isolates provides evidence for multiple genotypes
	JOURNAL	AIDS 7, 769-780 (1993)
	STANDARD COMMENT	Kindly provided prior to publication by Henry M. Jackson Foundation Research Laboratory, 1500 East Gude Drive, Rockville, MD 20850. The LBV23-10 gag sequence clusters with HIV-1 A subtype sequences. Location/Qualifiers 1..1459 /product="gag protein" /gene="gag" /codon_start=1 /translation="MGARASVLGGKLDISWEKILRPRGKKRYRLKHLVYASRELERF ALNPGGLETAEGCCQMLEOLOSALTGTSEELKSLFNTIALTMCVHGRIDVRKEALD KLEIEINRSKRKTQOAARATSSSONPVYONAGOMGTOPMSPRTANAVAYIEEKA FSFEVYIPMFALSLEGATPODLNMNLNVGHGOAOMLTKTINEEAAREPVPVHAHG PIPPCGMRBREGSDIAGTTSNLDOIGWMTSNRPITPEGLTYKMYIVLGKLKYRMISP VTLLDRQGKEPFRRDVDFRFRVLAEEASTQVEKMMEETLLVNANPCKSILURL GPAATLEEMWTACQGVGGPGHKARVALVAESASVOQTNIIMNRGNFRQARTIKFCNGK EGLRLAKCAKPARRCGWCKCEKHOKMDCTERANQENIGKIWPSSKGKPNFPOSREP TAPPASFSGGEETISSPKQEPEDKGIYLPRTLSKX"
	FEATURES CDS	



BASE COUNT 533 a 275 c 369 g 282 t  
ORIGIN gag cds start

Query Match  
Best Local Similarity 42.1%; Pred. No. 2.42e-01;  
Matches 33; Conservative 13; Mismatches 31; Indels 0; Gaps 0;

Db 269 aaagagatagctaaagacacacaggaagccttagataataatagagaaataaata 328  
167 ARAARTNGAYCTNACNAATTAARGAGCNYTGARACNTTYGNGNGTNAARMGM 226

Db 329 agagcaacaagaagac 344  
227 GNTTYAAYGARACNAC 242

LOCUS 8 HIV1415 1459 bp ss-RNA VRL 20-APR-1993  
DEFINITION Human immunodeficiency virus type 1, isolate V1415 taken from a Rwandan national residing in Belgium, gag region.  
ACCESSION L11791  
SOURCE Human immunodeficiency virus type 1 (HIV-1), Rwandan isolate V1415.  
ORGANISM  
Virus; ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
REFERENCE 1 (bases 1 to 1459)  
Louwagie, J.J., McCutchan, F., Brennan, T., Peeters, M., Brennan, T., Sanders-Buell, E., Eddy, G., van der Groen, G., Fransen, K., Gershby-Damet, M., Deleys, R. and Burke, D.  
Phylogenetic analysis of gag genes from seventy international HIV-1 isolates provides evidence for multiple genotypes  
AIDS 7, 769-780 (1993)

JOURNAL  
STANDARD  
COMMENT Kindly provided prior to publication by Henry M. Jackson Foundation Research Laboratory, 1500 East Gude Drive, Rockville, MD 20850.  
The V1415 gag sequence clusters with HIV-1 A subtype sequences.  
LOCATION/Qualifiers  
1..1459  
/product="gag protein"  
/gene="gag"  
/codon\_start=1  
/translation="MGARASYLSGGKLDAMERIKLRPGKKRYKHLWASRELEDR  
ALNGLLETAAGCOILEQLOPALKTGTEELKSYNTVATLYCVHORIDVDTAEALD  
KIEIKNNKNGKTOQAAGTGNSSVSONPIYVNAQOGPHQHSIPRTLNAMRVYE  
EKASPEVYIPMFSALESGATPGDNLNMIYVGGQAOMKDIINEEAAMWDLHPV  
HAGRIPIPGOMREPGSDIAGTSTTQEOIAMNTGNPPVVDIYKRIITILNKIVRA  
YSPSIDLYKQKPKPEPDYDFRFTKILRAQATQEVAGWMTETLLVONANPDCKTIL  
RALGTGATLEEMNTACQGVGPGHAKAVLAEMSOVNTIMORGNGKGORIKCFN  
CGKEGHLARNCRAPRKCGKCGEGHQMCKDTEQANFLCKIWSSSKGRGPNFPOS  
PEPTAPPAEIVMGSEISPPKQEQREQAPPLVLSKX"  
1..1459  
/organism="Human immunodeficiency virus type 1"  
/isolate="V1415"  
/cell\_type="lymphocyte"  
/proviral  
/sequenced\_mol="DNA"  
/tissue\_type="blood"

BASE COUNT 542 a 278 c 366 g 273 t  
ORIGIN gag cds start

Query Match  
Best Local Similarity 4.7%; Score 126; DB 3; Length 1459;  
Matches 33; Conservative 13; Mismatches 31; Indels 0; Gaps 0;

Db 269 agagatagctaaagacacacaggaagccttagataataatagagaaataaata 328  
167 ARAARTNGAYCTNACNAATTAARGAGCNYTGARACNTTYGNGNGTNAARMGM 226

Db 329 agagcaacaagaagac 344  
227 GNTTYAAYGARACNAC 242

JOURNAL  
STANDARD  
COMMENT Kindly provided prior to publication by Henry M. Jackson Foundation Research Laboratory, 1500 East Gude Drive, Rockville, MD 20850.  
The V1415 gag sequence clusters with HIV-1 A subtype sequences.  
LOCATION/Qualifiers  
1..1459  
/product="gag protein"  
/gene="gag"  
/codon\_start=1  
/translation="MGARASYLSGGKLDAMERIKLRPGKKRYKHLWASRELEDR  
ALNGLLETAAGCOILEQLOPALKTGTEELKSYNTVATLYCVHORIDVDTAEALD  
KIEIKNNKNGKTOQAAGTGNSSVSONPIYVNAQOGPHQHSIPRTLNAMRVYE  
EKASPEVYIPMFSALESGATPGDNLNMIYVGGQAOMKDIINEEAAMWDLHPV  
HAGRIPIPGOMREPGSDIAGTSTTQEOIAMNTGNPPVVDIYKRIITILNKIVRA  
YSPSIDLYKQKPKPEPDYDFRFTKILRAQATQEVAGWMTETLLVONANPDCKTIL  
RALGTGATLEEMNTACQGVGPGHAKAVLAEMSOVNTIMORGNGKGORIKCFN  
CGKEGHLARNCRAPRKCGKCGEGHQMCKDTEQANFLCKIWSSSKGRGPNFPOS  
PEPTAPPAEIVMGSEISPPKQEQREQAPPLVLSKX"

RESULT 9  
LOCUS HIV1557 1468 bp ss-RNA VRL 20-APR-1993  
DEFINITION Human immunodeficiency virus type 1, Zairean isolate V1557, gag region.  
ACCESSION L11793  
SOURCE Human immunodeficiency virus type 1 (HIV-1), Zairean isolate V1557.  
ORGANISM  
Human immunodeficiency virus type 1  
Virus; ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
REFERENCE 1 (bases 1 to 1468)  
Louwagie, J.J., McCutchan, F., Brennan, T., Peeters, M., Brennan, T., Sanders-Buell, E., Eddy, G., van der Groen, G., Fransen, K., Gershby-Damet, M., Deleys, R. and Burke, D.  
Phylogenetic analysis of gag genes from seventy international HIV-1 isolates provides evidence for multiple genotypes  
AIDS 7, 769-780 (1993)

JOURNAL  
STANDARD  
COMMENT Kindly provided prior to publication by Henry M. Jackson Foundation Research Laboratory, 1500 East Gude Drive, Rockville, MD 20850.  
The authors group the outlying V1557 gag sequence with V1525 as subtype G.  
LOCATION/Qualifiers  
1..1468  
/product="gag protein"  
/gene="gag"  
/codon\_start=1  
/translation="MGARASYLSGGKLDAMERIKLRPGKKRYKHLWASRELEDR  
ALNGLLETAAGCOILEQLOPALKTGTEELKSYNTVATLYCVHORIDVDTAEALD  
KIEIKNNKNGKTOQAAGTGNSSVSONPIYVNAQOGPHQHSIPRTLNAMRVYE  
EKASPEVYIPMFSALESGATPGDNLNMIYVGGQAOMKDIINEEAAMWDLHPV  
HAGRIPIPGOMREPGSDIAGTSTTQEOIAMNTGNPPVVDIYKRIITILNKIVRA  
YSPSIDLYKQKPKPEPDYDFRFTKILRAQATQEVAGWMTETLLVONANPDCKTIL  
RALGTGATLEEMNTACQGVGPGHAKAVLAEMSOVNTIMORGNGKGORIKCFN  
CGKEGHLARNCRAPRKCGKCGEGHQMCKDTEQANFLCKIWSSSKGRGPNFPOS  
PEPTAPPAEIVMGSEISPPKQEQREQAPPLVLSKX"  
1..1468  
/organism="Human immunodeficiency virus type 1"  
/isolate="V1557"  
/cell\_type="lymphocyte"  
/proviral  
/sequenced\_mol="DNA"  
/tissue\_type="blood"

BASE COUNT 538 a 277 c 363 g 290 t  
ORIGIN gag cds start

Query Match  
Best Local Similarity 4.6%; Score 122; DB 3; Length 1468;  
Matches 27; Conservative 13; Mismatches 17; Indels 1; Gaps 1;

Db 265 cttatgcgtacataagaat-agaatgaagaagacacaggaagccttagagaagt 311  
152 CNAHTWSNTAYTTNGARARARYTNGAYTNACNAATTAARGAGCNYTGARACNTT 209

RESULT 10  
LOCUS HIVPH153 1492 bp ss-RNA VRL 20-APR-1993  
DEFINITION Human immunodeficiency virus type 1, Philippines isolate PH153, gag region.  
ACCESSION L11781  
SOURCE Human immunodeficiency virus type 1 (HIV-1), Philippines isolate PH153.  
ORGANISM  
Human immunodeficiency virus type 1  
Virus; ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
REFERENCE 1 (bases 1 to 1492)  
Louwagie, J.J., McCutchan, F., Brennan, T., Peeters, M., Brennan, T., Sanders-Buell, E., Eddy, G., van der Groen, G., Fransen, K., Gershby-Damet, M., Deleys, R. and Burke, D.  
Phylogenetic analysis of gag genes from seventy international HIV-1 isolates provides evidence for multiple genotypes  
AIDS 7, 769-780 (1993)

JOURNAL  
STANDARD  
COMMENT Kindly provided prior to publication by Henry M. Jackson Foundation



ORGANISM	Human immunodeficiency virus type 1
Viridae: ss-RNA enveloped viruses; Positive strand RNA virus; Retroviridae; Lentivirinae.	
1 (bases 7005 to 7514)	
Li, Y., Kappe, J. C., Conway, J. A., Price, R. W., Shaw, G. M. and	
REFERENCE	
AUTHORS	







```

CDS      <2092..5103
         /note="pol polypeptide (NH2-terminus uncertain; AA at
         2092)"
         /codon_start=2092
         join(5840..6054,8384..8474)
         /note="lat protein"
         /codon_start=5840
         join(5979..6054,8384..8658)
         /note="rev protein"
         /codon_start=5979
         2..635
         /note="5' LTR"
         332..332
         /note="g in 7A,9B; a in 8A"
         340..340
         /note="g in 7A,9B; a in 8A"
         378..387
         /note="Spl binding site III"
         389..398
         /note="Spl binding site II"
         400..409
         /note="Spl binding site I"
         454..551
         /note="R repeat 5' copy"
         /note="R repeat 5' copy"
         455..9652
         /note="genomic mRNA"
         455..9652
         /note="tat, rev, nef subgenomic mRNA"
         637..654
         /note="primer (Lys-tRNA) binding site"
         745..5783
         /note="tat, rev, nef subgenomic mRNA intron 1"
         791..2299
         /note="gag polypeptide"
         /codon_start=791
         4233..4233
         /note="a in 8A; g in 8B"
         4677..4677
         /note="t in 8A; c in 8B"
         5048..5626
         /note="vif protein"
         /codon_start=5048
         5566..5859
         /note="vpr protein"
         /codon_start=5566
         5840..6054
         /note="tat protein, exon 2 (first expressed exon)"
         5979..6054
         /note="rev protein, exon 2 (first expressed exon)"
         6055..8383
         /note="tat cds intron 2"
         6055..8383
         /note="rev cds intron 2"
         6055..8383
         /note="tat, rev, nef subgenomic mRNA intron 2"
         6071..6187
         /note="vpu protein"
         /codon_start=6071
         6215..6215
         /note="a in 7D; g in 7A,9B"
         6233..8800
         /note="envelope polypeptide"
         /codon_start=6233
         8384..8474
         /note="tat protein, exon 3 (AA at 8385)"
         8384..8658
         /note="rev protein, exon 3 (AA at 8386)"
         8802..9434
         /note="nef protein"
         /codon_start=8802
         9103..9736
         /note="3' LTR"
         9555..9652
         /repeat_region

```

```

misc_signal      /note="R repeat 3' copy"
                  9628..9633
                  /note="mRNA polyadenylation signal"
BASE COUNT      3445 a 1738 c 2177 t
ORIGIN          5' terminus of LTR-U3.
Query Match      4.5%; Score 121; DB 3; Length 9737;
Best Local Similarity 40.3%; Pred. No. 7.15e-01;
Matches 29; Conservative 14; Mismatches 29; Indels 0; Gaps 0;
Db 1059 aaagatagatgtaaaagaccacgaagagctttagaagaatagaggaagcaaaaca 1118
QY 167 ARAARYTNGAYGTNACNAAYATHAARGRCGNTYGGNGGCTNAAARGNM 226
Db 1119 aaagtaagaaa 1130
QY 227 GNTTYAAYGARA 238

```

Search completed: Sat Nov 27 10:18:33 1999  
 Job time : 8 secs.



\*\*\*\*\*  
 Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd  
 \*\*\*\*\*

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_tpn n.a. - n.a. Smith-Waterman search, using a protein query  
 which has been backtranslated into n.a. using IUPAC symbols

Run on: Sat Nov 27 10:20:03 1999; Maspar time 159.66 Seconds  
 865.603 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-4  
 Description: (1-215) from US09103287.pep  
 Perfect Score: 2663  
 N.A. Sequence: 1 TTYAARGAVYNSGAYGAYAT.....TNGMNTGARAAVGCNTTY 645  
 Comp: AARTTYTRMSNCTRCRTA.....ANCCNACTTITTCGNAAR

Scoring table: TABLE bktranslates2  
 Gap 30

Nmatch STD : Dbase 0; Query 0

Searched: 271905 segs, 107135622 bases x 2

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database:

n-genesegs  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21 22:part22 23:part23  
 24:part24 25:part25 26:part26 27:part27 28:part28  
 29:part29 30:part30 31:part31 32:part32 33:part33  
 34:part34 35:part35 36:part36 37:part37 38:part38  
 39:part39 40:part40 41:part41 42:part42 43:part43  
 44:part44 45:part45 46:part46 47:part47 48:part48  
 49:part49 50:part50 51:part51 52:part52 53:part53  
 54:part54 55:part55 56:part56 57:part57 58:part58  
 59:part59 60:part60

Statistics: Mean 50.914; Variance 239.321; scale 0.213

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	2573	96.6	2424	58	V4703	Staphylococcus aureus
2	2557	96.0	660	60	V80065	Partial nucleotide se
3	2557	96.0	1351	60	V99650	UDP-N-acetylmutaric
4	2329	87.5	619	48	V53479	DNA encoding a Staphy
5	934	35.1	677	60	X13717	Enterococcus faecalis
6	749	28.1	1267	46	V27381	Streptococcus pneumonia
7	749	28.1	11864	47	V52194	Streptococcus pneumonia
8	699	26.2	1825	47	V43027	Streptococcus pneumonia

Result ID	Score	Query Match	Length	DB ID	Description	Pred. No.
9	222	8.3	861	30	T77403	H. pylori cytoplasmic
10	222	8.3	861	30	T67723	H. pylori cytoplasmic
11	202	7.6	1365	30	T67859	H. pylori cytoplasmic
12	183	6.9	843	47	V52489	Streptococcus pneumonia
13	183	6.9	1003	46	V27378	Streptococcus pneumonia
14	183	6.9	1350	36	T89154	Streptococcus pneumonia
15	183	6.9	1350	52	V33971	S. pneumoniae Mard mR
16	183	6.9	1353	52	V33970	S. pneumoniae Mard mR
17	183	6.9	6217	50	V65252	DNA encoding S. pneum
18	183	6.9	6693	47	V52328	Streptococcus pneumonia
19	152	5.7	534	52	V59408	Degenerate nucleotide
20	151	5.7	1532	47	V45173	Synthetic HIV gag gen
21	140	5.3	501	3	N50026	Sequence encoding new
22	140	5.3	501	3	N50024	Sequence encoding new
23	141	5.3	1311	59	V81395	Degenerate human tumo
24	136	5.2	2193	60	V84038	Human gelsolin gene c
25	139	5.2	8247	2	N70996	Sequence encoding N-t
26	135	5.1	1026	30	T68253	H. pylori cytoplasmic
27	135	5.1	1071	48	V19280	Pseudomonas fluoresce
28	137	5.1	1782	1	Q03856	Heat shock protein (H
29	133	5.0	501	3	N50024	Sequence encoding new
30	134	5.0	510	42	V28775	Human interleukin-xx
31	134	5.0	984	17	Q94336	Degenerate Alteromona
32	132	5.0	1640	2	N60590	Sequence encoding mat
33	132	5.0	1641	17	T08561	Human CSF-1 cDNA clon
34	132	5.0	1641	60	X13057	Enterococcus faecalis
35	132	5.0	1642	2	N80775	Sequence of a cDNA en
36	132	5.0	1642	1	N91084	Colony stimulating fa
37	132	5.0	1642	37	V01080	Gene for human colony
38	132	5.0	1642	33	T80591	Human colony stimulat
39	132	5.0	1642	26	T58753	Human short form CSF-
40	132	5.0	1880	46	V35031	Arabidopsis thaliana
41	132	5.0	2302	35	T94747	Human colony stimulat
42	132	5.0	2303	13	Q73786	Partial PCR fragment
43	132	5.0	2500	7	Q43710	Sequence of clone lam
44	132	5.0	4315	60	X12938	Enterococcus faecalis
45	134	5.0	9767	60	X13067	Enterococcus faecalis

## ALIGNMENTS

RESULT 1  
 ID V74703 standard; DNA; 2424 BP.  
 AC V74703:  
 DE 16-MAR-1999 (first entry)  
 ST Staphylococcus aureus contig SED ID #392.  
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.  
 OS Staphylococcus aureus.  
 FH key location/Qualifiers  
 FT misc-feature 1141..1200  
 FT /tag="a"  
 FT /note="these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

EP-766519-A2.  
 30-JUL-1997.  
 PF 07-JAN-1997: 100117.  
 PR 05-JAN-1996: US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
 PI Rosen CA.  
 DR NPI: 97-374922/35.  
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines  
 PS Claim 1; Page 1287-1288; 3271pp; English.  
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 of the invention. The DNA sequences are recorded on a computer readable  
 medium, preferably selected from a floppy or hard disk, random access







Oy	421	CAYMNGNTTYYTNTGICARARHTHTTYGGWMSNATHMGNCARAAHYACNGNGCNTYTNACN	480
Db	482	atacaagaatttaattgataaaaattggagglygcactcgtcttcatlaatgaagacttatat	541
Oy	481	ATHCARGAAYTTNATHCAVAAARAATHGARAGCGCWMSYNTNATHAAYGARGAWSNATHAAY	540
Db	542	gtatagaacaatttgataatactgtctgtttttattatagggtgcaggtgatattcaaaaa	601
Oy	541	GTTYNYGARCACATTGYAVAAAYGCGNGTCGTNTYTNTYATGTGGCNGCGNGANGYATHCARAR	600
Db	602	ttaacaatatcatattagataaaattagcatcataaaatcgcttt	646
Oy	601	YNMCARAAATGCNTATYTNCAIYARNTNGGNATCAARAATGCNTTY	645
RESULT	3		
ID	V99650	standard; DNA; 1351 BP.	
AC	V99650:		
DT	17-MAR-1999	(first entry)	
DE	UDP-N-acetylmuramate:L-alanine ligase (MurC polypeptide) encoding DNA.		
KW	MurC gene; UDP-N-acetylmuramate:L-alanine ligase; MurC polypeptide;		
RW	bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;		
OS	immunogen; drug; genetic immunisation; ds.		
ST	Staphylococcus aureus.		
FH	Key	Location/Qualifiers	
FT	CD5	22..1335	
FT	/tag=	"a"	
FT	/gene=	"MurC"	
FT	/product=	"UDP-N-acetylmuramate:L-alanine ligase (MurC polypeptide)"	
PN	EP-889123-A2.		
PD	07-JAN-1999.		
PE	26-JUN-1998; 305064.		
PR	03-JUL-1997; US-052720.		
PA	(SMIK ) SMIITHKLINE BECHAM CORP.		
P1	(SMIK ) SMIITHKLINE BECHAM PLC.		
PI	Burnham MKR, Wallis NG;		
DR	WPI; 99-062655/06.		
P-PSDB:	W87771.		
PT	New isolated MurC polypeptide from Staphylococcus aureus and related nucleic acid - useful in diagnosis, treatment and prevention of bacterial infections		
PT	Claim 2; Pages 3-4; 39pp; English.		
CC	The present sequence represents a MurC gene encoding a Staphylococcus aureus UDP-N-acetylmuramate:L-alanine ligase (MurC polypeptide). Host cell containing an expression system comprising the MurC gene can be used for the recombinant production of the polypeptide. Agonists or the MurC polypeptide are used to treat conditions requiring increased activity or expression of the polypeptide. Antagonists, inhibitory nucleic acid or competitive polypeptide are useful for inhibiting the polypeptide e.g. bacteriacial (especially S. aureus) infections. They are also useful against Helicobacter pylori infections and related cancers, ulcers and gastritis. The antibacterial agents are useful to treat in-dwelling devices for infection prevention or generally as wound treatments to prevent adhesion of bacteria to matrix proteins. The MurC polypeptide is also useful for diagnosing or prognosing a (susceptibility to) disease, for raising antibodies: to identify modulators or specific receptors; in rational drug design and as an immunogen for vaccines. The MurC gene sequences are useful in antisense/ribozyme therapeutics: to detect mutant MurC gene; for chromosomal mapping; to determine bacterial serotype; and for genetic immunisation.		
SC	Sequence	1351 BP; 471 A; 166 C; 260 G; 454 T;	
Query Match	96.0%; Score 2557; DB 60; Length 1351;		
Best Local Similarity	59.4%; Pred. No. 9,99e-211;		
Matches	383; Conservative 157; Mismatches 105; Indels 0; Gaps 0;		
Db	688	tttaagaatttcgataacattatgctcaaaatcatcaaatccgataaagttaactgct	747
Oy	1	TYYAARGATWSNCAIYAATHTYATGYGCNCARATHTTTCARATHACNGAYTAARGNACNGCN	60
Db	748	tttgatgltatgtggaatggtgagtttattagatcacttcctgtctccacaatalgtygac	807

QY	61	GTNCAIGTNTATYGTNGAYGNGNARKRTTYTAYGAYCATTTYYTMMSCCNCAKRTAYGNGNAY	120
Db	808	catacagtttaaatgcatagctgtaattgcatgattagattcagaagcgtatgctt	867
QY	121	CAYACNGNTYNNAYGCGNTYNGCGNTGNATHGCMATHMSTATYTTNGARPARRTNCAIYGTN	180
Db	868	acaaatltaaagaagcattagaagaagcttgtygtggtttaaagctgcgttcattgaanaat	927
QY	181	ACNAAVYTHARGARGCGNTYNGARACNTTYGGNGGNGTNAARMGMNGNTTYAGARCN	240
Db	928	acaaatgcgaatcaagttatcttgatgattatgacacaccatccaagagaanaatgctc	987
QY	241	ACNATHGCMNAYCARGTGNATHGTNGAYGAYTAAGCCAYCCAMWGNGARATHMSNGCN	300
Db	988	acaaatgaccagacgaagaagaattccacataaagaagtgtgtgcgatttcaacca	1047
QY	301	ACNNTHGAYACNGCMNMNAAARATRYCCNCAVAAAGAGTNGTNGCMTNTTYARCCN	360
Db	1048	cacacttcctcagaacacagaagcattttaaagtattgctgcgaagattgatgaagca	1107
QY	361	CAYACNTTYSNMNMACNCARCGCNTTYTNAAGAATTTGCCGARKSNYTNMSSNARCGN	420
Db	1108	gacgtgtatctcatgattggaattttggcctcatagagaanaatctgcgcattacg	1167
QY	421	GAYMGNGNTTYNTNYOYGARATHHTTYGGMWSNATHMGNSARAAACNGNCGCNTNTMACN	480
Db	1168	atacagaatttaattgataaattggagtgctacgctgttcattaaatgaatcattlaat	1227
QY	481	ATHCARGAYYYNATHGAYAAARATHHGARGCGCMWSNYTNATHAAGARGAYMSNTAAHAY	540
Db	1228	gtattagaacaattgataatgctgtgttttatttaatgtygtcagtyuatctcaana	1287
QY	541	GTNYTNGARCARTTYGAYAYGCGNGTNGTYNTTYATGCGMCGNGGNAVYATHCARAAR	600
Db	1288	ttacaaatgcatattagatlaaataggcatgacatgaaanaatgcgttt	1332
QY	601	YTNCAARAAYGCGNTATYYTNGAYAYTNGNATTAATAAAAGCMTTY	645

RESULT 4  
ID V53479 standard: DNA; 619 BP.  
AC V53479;  
DT 30-Oct-1998 (first entry)  
DE DNA encoding a Staphylococcus aureus protein of unknown function.  
KW Staphylococcus aureus protein; immune response induction; eye infection;  
KW antibody production; T-cell immune response; gastrointestinal infection;  
KW respiratory infection; inhibitor; bacterial infection; cardiac infection;  
KW central nervous system; kidney infection; urinary tract infection;  
KW antimicrobial compound identification; broad spectrum antibiotic;  
KW therapy; ss.  
OS Staphylococcus aureus.  
PN EP-841394-A2.  
PD 13-MAY-1998.  
PP 24-SEP-1997; 307485.  
PR 24-SEP-1996; US-027032.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PI Black MT, Burnham MKR, Hodgson JE, Knowles DJC,  
PI Llorente MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,  
PI Ward JM;  
DR WPI: 98-252940/23.  
DR P-PSDB: W77686.  
PT New nucleic acid sequences from Staphylococcus aureus WCHU29 -  
PT useful in vaccines and for treatment of bacterial infections of e.g.  
PT respiratory tract and central nervous system  
PS Claim 1; Page 146; 390PP: English.  
CC This sequence encodes a Staphylococcus aureus protein of unknown  
CC function, and represents a DNA sequence of the invention.  
CC The DNA sequences were isolated from Staphylococcus aureus WCHU29  
CC (NCIMB 40771). Host cells containing the DNA sequences are used to  
CC produce polypeptides or fragments. The proteins are used in the treatment  
CC of disease, for inducing an immune response by administering them, to















```

Query Match      8.3%: Score 222: DB 30: Length 861:
Best Local Similarity 36.4%: Pred. No. 2.48e-05:
Matches      60: Conservative      34: Mismatches 71: Indels 0: Gaps 0:

Db      353 taagcgcttagaagaattaattagaagaattagaataattatgaatttaaa 412
      155 THWSTATVYTGGAARARTNGAYGTNMCNMAVAFHARARGRCVYTGARACNTTGGNG 214
Qy      215 GNGTNAAMGMGMGNTTAAVYGARACNACNMTHGCNMAVCAKGTWATGTTGAGATYATG 274
Db      413 gcatataaaacgcttagattttgcacaaataacatccatcattgatgatgaag 472
      215 GNGTNAAMGMGMGNTTAAVYGARACNACNMTHGCNMAVCAKGTWATGTTGAGATYATG 274
Qy      473 cccgccaccctgaactgagcgcaccttaaaagcgtaga 517
      275 CACAATCAATCANNMGARATHWSSNCCNACNATHGAYACNCGMNA 319

RESULT      10
ID      T67723 standard: DNA; 861 BP.
AC      T67723:
DE      28-JUL-1997 (first entry)
DEF     H. pylori cytoplasmic protein ORF 11253.aa.
KW      Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW      identification; binding compound; bacterium; life cycle; activator;
KW      bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW      diagnosis; ds.
OS      Helicobacter pylori.
FH      Key
FT      Location/Qualifiers
FT      cds
FT      1..861
FT      /*tag= a
FT      /transl_except= (pos: 4..6, aa: Xaa)
FT      /transl_except= (pos: 397..399, aa: Xaa)
FT      /transl_except= (pos: 496..498, aa: Xaa)
FT      /note= "Xaa = Unknown"

PN      W09640893-A1.
PR      19-DEC-1996.
PR      06-JUN-1996: U09122.
PR      07-JUN-1995: US-487032.
PR      01-APR-1996: US-630405.
PI      (ASTR) ASTRA AB.
PI      Berglinch OT, Smith D, Mellgaerd BL;
DR      WPI: 97-052306/05.
DR      P-RSDS: W20102.

PT      Helicobacter pylori nucleic acid sequences and related
PT      polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT      infection, and to detect Helicobacter
PS      Claim 9; Page -, 1481pp; English.
SC      This sequence encodes a H. pylori cytoplasmic protein involved in
CC      outer membrane or cell wall biosynthesis.
CC      The protein may be used in a vaccine to prevent or treat H. pylori
CC      infection or to identify H. pylori polypeptide binding compounds,
CC      useful as potential H. pylori life cycle activators or inhibitors.
CC      The genomic sequence of H. pylori (ATCC 55679) was determined from
CC      overlapping configs generated by mechanically shearing the bacterial
CC      DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC      and the predicted coding regions defined by computer evaluation. To
CC      identify likely H. pylori antigens for vaccine development, the amino
CC      acid sequences predicted from various ORF were analysed for significant
CC      homology to other known or exported membrane proteins. Having identified
CC      and determined the sequences of interest, particular regions can be
CC      isolated from H. pylori by PCR amplification for recombinant polypeptide
CC      production, e.g. in E. coli hosts.
CC      Note: This DNA sequence is not reproduced in the specification and
CC      has been derived from the related specification, W09718098.
SQ      Sequence      861 BP;      291 A;      130 C;      169 G;      268 T;

Query Match      8.3%: Score 222: DB 30: Length 861:
Best Local Similarity 36.4%: Pred. No. 2.48e-05:
Matches      60: Conservative      34: Mismatches 71: Indels 0: Gaps 0:

Db      353 taagcgcttagaagaattaattagaagaattagaataattatgaatttaaa 412
      155 THWSTATVYTGGAARARTNGAYGTNMCNMAVAFHARARGRCVYTGARACNTTGGNG 214
Qy      215 GNGTNAAMGMGMGNTTAAVYGARACNACNMTHGCNMAVCAKGTWATGTTGAGATYATG 274
Db      413 gcatataaaacgcttagattttgcacaaataacatccatcattgatgatgaag 472
      215 GNGTNAAMGMGMGNTTAAVYGARACNACNMTHGCNMAVCAKGTWATGTTGAGATYATG 274
Qy      473 cccgccaccctgaactgagcgcaccttaaaagcgtaga 517
      275 CACAATCAATCANNMGARATHWSSNCCNACNATHGAYACNCGMNA 319

```

Oy	155	THWSTAYVTNGAARARTYNTGATGNACNAAVATHHARGARGCVYTGADACNTTGGNG	214
Dd	413	gcattaaacagcgttgatatttcgcaaaaacaatccattctcatgatgaatg	472
Oy	215	GNGTNAAAMGMNGMTTYAAAGARACNNACNMTHGCAACGATGTATGTGACGATYAG	274
Dd	473	ccccaccacctgaattggcccactttaaaggcgtaga	517
Oy	275	CNCATCAYCCNMNGARHTWSNCGNCACNMTHGAVACNCGMNA	319

RESULT 11					
DD	T67859	standard; DNA;	1365 BP.		
AC	T67859;				
DT	14-JUL-1997	(first entry)			
KW	H. pylori cytoplasmic protein ORF 01epj0520orf27.				
DV	Vaccine; prevention; treatment; infection; identification;				
KW	binding compound; bacterium; life cycle; activator; bacteria;				
KW	inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;				
KW	cytoplasmic; outer membrane; cell wall; biosynthesis; ds.				
OS	Helicobacter pylori.				
FH	Key	Location/Qualifiers			
FT	cds	1..1365			
FT		/tag= a			
FT		/note= "no stop codon given"			
PN	W09640893.A1.				
PD	19-DEC-1996.				
PF	06-JUN-1996; U09122.				
PR	07-JUN-1995; US-487032.				
PR	01-APR-1996; US-630405.				
PA	(ASTR) ASTRA AB.				
P1	Berglindh OT; Smith D, Mellgaard BL;				
DR	WPJ; 97-052306/05.				
DR	P-PSDB; W20606.				
PT	Helicobacter pylori nucleic acid sequences and related				
PT	polypeptide(s) -- useful for vaccines to treat or prevent H. pylori				
PT	infection, and to detect Helicobacter				
PS	Claim 9; Page 752; 1481pp; English.				
CC	The present sequence encodes a Helicobacter pylori cytoplasmic				
CC	protein involved outer membrane or cell wall biosynthesis.				
CC	The protein may be used in a vaccine to prevent or treat H. pylori				
CC	infection or to identify H. pylori polypeptide binding compounds,				
CC	useful as potential H. pylori life cycle activators or inhibitors.				
CC	The genomic sequence of H. pylori (ATCC 55679) was determined from				
CC	overlapping contigs generated by mechanically shearing the				
CC	bacterial DNA. The sequences were analysed for ORF of at least 180				
CC	nucleotides, and the predicted coding regions defined by computer				
CC	evaluation. To identify likely H. pylori antigens for vaccine				
CC	development, the amino acid sequences predicted from various ORF				
CC	were analysed for significant homology to other known or exported				
CC	membrane proteins. Having identified and determined the sequences				
CC	of interest, particular regions can be isolated from H. pylori by				
CC	PCR amplification for recombinant polypeptide production, e.g. in				
CC	E. coli hosts.				
SQ	Sequence 1365 BP; 444 A; 231 C; 288 G; 402 T;				

Query Match 7.6%; Score 202; DB 30; Length 1365;					
Best Local Similarity 36.4%; Pred. No. 7.08e-04;					
Matches	60;	Conservative	32;	Mismatches	73;
				Indels	0;
				Gaps	0;

Dd	860	taagcgctttagaatataattagaagaataatagaataattatgaatttaaag	919
Oy	155	THWSTAYVTNGAARARTYNTGATGNACNAAVATHHARGARGCVYTGADACNTTGGNG	214
Dd	920	gcattaaaaaacgcttgatatatttcgcaaaaacaagatccattctcatgatgaatg	979
Oy	215	GNGTNAAAMGMNGMTTYAAAGARACNNACNMTHGCAACGATGTATGTGACGATYAG	274
Dd	980	ccccaccacctgaattggcccactttaaaggcgtaga	1024
Oy	275	CNCATCAYCCNMNGARHTWSNCGNCACNMTHGAVACNCGMNA	319











**THIS PAGE BLANK (USPTO)**



\*\*\*\*\*  
W O R L D  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msrch\_tpn n.a. - n.a. Smith-Waterman search, using a protein query  
which has been backtranslated into n.a. using IUPAC symbols

Run on: Sat Nov 27 10:18:50 1999; Maspar time 49.83 Seconds  
1119,486 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-4  
Description: (1-215) from US09103287.pep  
Perfect score: 2663  
N.A. Sequence: 1 TTYAARGAYWSNGAYGAYAT.....TNGNATGARAAGONTTY 645  
Comp: AARTTYCTFMSNCTRCRTTRA.....ANCMNACTTYYTTRCGNAR

Scoring table: TABLE bktranslate2  
Gap 30

Nmatch STD : Dbase 0; Query 0

Searched: 165359 seqs, 43243793 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

n-issued  
1:3A\_COMB 2:5E\_COMB 3:5C\_COMB 4:PCT9\_COMB 5:backfiles1

Statistics: Mean 47.980; Variance 191.666; scale 0.250  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	268	10.1	7218	2	US-08-232-	Sequence 14, Applicatio	9.12e-12
2	183	6.9	1350	2	US-08-665-	Sequence 3, Applicatio	2.60e-04
3	183	6.9	1350	2	US-08-665-	Sequence 1, Applicatio	2.60e-04
4	183	6.9	1350	3	US-08-843-	Sequence 3, Applicatio	2.60e-04
5	183	6.9	1353	3	US-08-843-	Sequence 1, Applicatio	2.60e-04
6	138	5.2	2588	4	PCT-US91-0	Sequence 1, Applicatio	1.00e+00
7	138	5.2	2608	1	US-07-725-	Sequence 1, Applicatio	1.00e+00
8	137	5.1	1760	2	US-08-413-	Sequence 118, Applicat	1.19e+00
9	132	5.0	1642	2	US-08-426-	Sequence 3, Applicatio	2.82e+00
10	132	5.0	1642	2	US-08-401-	Sequence 3, Applicatio	2.82e+00
11	132	5.0	1642	1	US-08-354-	Sequence 3, Applicatio	2.82e+00
12	132	5.0	1642	1	US-07-999-	Sequence 3, Applicatio	2.82e+00
13	132	5.0	1642	2	US-08-426-	Sequence 3, Applicatio	2.82e+00
14	132	5.0	1659	1	US-08-231-	Sequence 2, Applicatio	2.82e+00
15	132	5.0	1659	1	US-08-231-	Sequence 1, Applicatio	2.82e+00
16	132	5.0	1659	1	US-08-231-	Sequence 2, Applicatio	2.82e+00
17	132	5.0	1659	1	US-08-231-	Sequence 1, Applicatio	2.82e+00
18	132	5.0	2173	1	US-08-036-	Sequence 6, Applicatio	2.82e+00
19	132	5.0	2302	1	US-07-999-	Sequence 1, Applicatio	2.82e+00

20	132	5.0	2302	1	US-08-426-	Sequence 1, Applicatio	2.82e+00
21	132	5.0	2302	1	US-08-354-	Sequence 1, Applicatio	2.82e+00
22	132	5.0	2302	2	US-08-401-	Sequence 1, Applicatio	2.82e+00
23	132	5.0	2302	2	US-08-426-	Sequence 1, Applicatio	2.82e+00
24	132	5.0	2309	1	US-08-036-	Sequence 10, Applicati	2.82e+00
25	132	5.0	2692	1	US-08-036-	Sequence 14, Applicati	2.82e+00
26	132	5.0	3719	3	US-08-920-	Sequence 10, Applicati	2.82e+00
27	132	5.0	3719	3	US-08-920-	Sequence 10, Applicati	2.82e+00
28	132	5.0	3719	3	US-08-921-	Sequence 10, Applicati	2.82e+00
29	132	5.0	3719	3	US-08-920-	Sequence 10, Applicati	2.82e+00
30	132	5.0	3719	3	US-08-362-	Sequence 10, Applicati	2.82e+00
31	132	5.0	3973	1	US-08-036-	Sequence 21, Applicati	2.82e+00
32	127	4.8	1428	4	PCT-US94-0	Sequence 11, Applicati	6.55e+00
33	127	4.8	1515	4	PCT-US94-0	Sequence 15, Applicati	6.55e+00
34	126	4.7	2277	3	US-08-676-	Sequence 2, Applicatio	7.74e+00
35	126	4.7	2277	2	US-08-676-	Sequence 2, Applicatio	7.74e+00
36	123	4.6	3033	1	US-08-095-	Sequence 3, Applicatio	1.27e+01
37	123	4.6	3033	2	US-08-480-	Sequence 3, Applicatio	1.27e+01
38	122	4.6	4276	2	US-08-343-	Sequence 3, Applicatio	1.50e+01
39	122	4.6	4276	1	US-07-973-	Sequence 3, Applicatio	1.50e+01
40	122	4.5	7785	3	US-08-276-	Sequence 1, Applicatio	1.50e+01
41	121	4.5	1367	4	PCT-US96-0	Sequence 3, Applicatio	1.77e+01
42	121	4.5	1367	3	US-08-948-	Sequence 3, Applicatio	1.77e+01
43	121	4.5	1367	2	US-08-453-	Sequence 3, Applicatio	1.77e+01
44	121	4.5	2028	4	PCT-US96-0	Sequence 1, Applicatio	1.77e+01
45	121	4.5	2028	2	US-08-453-	Sequence 1, Applicatio	1.77e+01

## ALIGNMENTS

RESULT 1  
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.  
AC xxxxxx  
DT  
DE Sequence 14, Application US/08232463  
CC Sequence 14, Application US/08232463  
CC Patent No. 5670367  
CC GENERAL INFORMATION:  
CC APPLICANT: DORNER, F.  
CC APPLICANT: SCHEIFLINGER, F.  
CC APPLICANT: FALKNER, F. G.  
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
CC NUMBER OF SEQUENCES: 52  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Foley & Lardner  
CC STREET: 1800 Diagonal Road, Suite 500  
CC CITY: Alexandria  
CC STATE: VA  
CC COUNTRY: USA  
CC ZIP: 22313-0299  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/232,463  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/935,313  
CC FILING DATE:  
CC APPLICATION NUMBER: EP 91 114 300.6  
CC FILING DATE: 26-AUG-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: BENT, Stephen A.  
CC REGISTRATION NUMBER: 29,768  
CC REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (703)836-9300  
CC TELEFAX: (703)833-4109  
CC TELEEX: 899149  
CC INFORMATION FOR SEQ ID NO: 14:



```

CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 7218 base pairs
CC      TYPE: nucleic acid
CC      STRANDEDNESS: single
CC      TOPOLOGY: linear
CC      IMMEDIATE SOURCE:
CC      CLONE: pTZ39c-Fls
CC      SEQUENCE 7218 bp, 1944 A, 1491 C, 1486 G, 1929 T, 368 OTHER.

CC      Query Match          10.1%: Score 268; DB 2; Length 7218;
CC      Best Local Similarity 7.4%: Pred. No. 9,12e-12;
CC      Matches 26; Conservative 172; Mismatches 180; Indels 0; Gaps

Cc      Db 1064 GATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1123
Cc      Cp 533 SWRICYTCRTDANANNSMNGCCCYTCDAATYTRCDATANARTCYGADTNGTNARN 474
Cc      Db 1124 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1183
Cc      Cp 473 GCNCGNCGRTTYTNCNDATNNSMCCAAATATTCGANAARRAANCNCRKNCGYTN 414
Cc      Db 1184 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1243
Cc      Cp 413 SMNARNSWTCNGCRAAYTCRTTNARRAANGCTGNGTCKNSMRANGTFRNGGYTGR 354
Cc      Db 1244 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1303
Cc      Cp 353 AANACNCGMNCMACYCTYTRTGNGGRTATYTTNCKNGCGTRTCADATNGTNCNSMD 294
Cc      Db 1304 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1363
Cc      Cp 293 ATYCNCKNGRGRTRTNGCARTATCTCNACACATNACTGRTNGCDATNGTNGTYTCR 234
Cc      Db 1364 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1423
Cc      Cp 233 TTRANCKNCMYTTNACNCCRAANGTYTCAARNGCYCTCTDATTGTNGTNCRCYN 174
Cc      Db 1424 YYYYYYYYYYGTGCAAT 1443
Cc      Cp 173 ARYTTYTCNARRTANSWDAT 154

Cc      RESULT          2
Cc      ID US-08-665-435A-3 STANDARD: DNA; UNC; 1350 BP.
Cc      AC xxxxxx
Cc      DT
Cc      DE Sequence 3, Application US/08665435A
Cc      CC Sequence 3, Application US/08665435A
Cc      CC Patent No. 5681694
Cc      CC GENERAL INFORMATION:
Cc      CC APPLICANT: Skatrud, Paul
Cc      CC APPLICANT: Peery, Robert
Cc      CC APPLICANT: Hoskins, Joann
Cc      CC APPLICANT: Wu, Chyun-yeh Earnest
Cc      CC TITLE OF INVENTION: Biosynthetic Gene Muid of Streptococcus
Cc      CC TITLE OF INVENTION: pneumoniae
Cc      CC NUMBER OF SEQUENCES: 3
Cc      CC CORRESPONDENCE ADDRESS:
Cc      CC ADDRESSEE: Eli Lilly and Company
Cc      CC STREET: Lilly Corporate Center
Cc      CC CITY: Indianapolis
Cc      CC STATE: Indiana
Cc      CC COUNTRY: US
Cc      CC ZIP: 46285
Cc      CC COMPUTER READABLE FORM:
Cc      CC MEDIUM TYPE: Floppy disk
Cc      CC COMPUTER: IBM PC compatible
Cc      CC OPERATING SYSTEM: PC-DOS/MS-DOS
Cc      CC SOFTWARE: PatentIn Release #1.0, Version #1.30
Cc      CC CURRENT APPLICATION DATA:
Cc      CC APPLICATION NUMBER: US/08/665,435A
Cc      CC FILING DATE:
Cc      CC CLASSIFICATION: 435

```

[illegible]







RESULT 6  
ID PCT-US91-02954-1 STANDARD; DNA; UNC; 2588 BP.  
AC xxxxxx  
DT  
DE Sequence 1, Application PC/TUS9102954  
CC Sequence 1, Application PC/TUS9102954  
CC GENERAL INFORMATION:  
CC APPLICANT: PEPINSKY, R. BLAKE  
CC APPLICANT: ROSA, MARGARET D.  
CC APPLICANT: STOSSEL, THOMAS P.  
CC TITLE OF INVENTION: MULTIMERIC GELSOLIN FUSION CONSTRUCTS  
CC NUMBER OF SEQUENCES: 14  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: FISH & NEAVE  
CC STREET: 875 Third Avenue  
CC CITY: New York  
CC STATE: New York  
CC COUNTRY: United States of America  
CC ZIP: 10022  
CC  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US91/02954  
CC FILING DATE: 19910503  
CC CLASSIFICATION: 436  
CC  
CC PRIORITY APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/520,368  
CC FILING DATE: 04-MAY-1990  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Haley Jr., James F.  
CC REGISTRATION NUMBER: 27,794  
CC TELEPHONE: (212) 715-0634  
CC TELEFAX: (212) 715-0634  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 715-0600  
CC TELEFAX: (212) 715-0634  
CC TELEX: 14-8367  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 2588 base pairs  
CC TYPE: NUCLEIC ACID  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC  
CC SEQUENCE 2588 BP: 550 A; 743 C; 800 G; 495 T; 0 OTHER.  
SQ  
Query Match 5.2%; Score 138; DB 4; Length 2588;  
Best Local Similarity 36.8%; Pred. No. 1.00e+00;  
Matches 32; Conservative 19; Mismatches 36; Indels 0; Gaps 0;

CC  
CC NUMBER OF SEQUENCES: 2  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
CC STREET: 1100 New York Avenue, Suite 600  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20005  
CC  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: IBM PC Compatible  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/725,083  
CC FILING DATE: 03-JUL-1991  
CC CLASSIFICATION: 435  
CC  
CC PRIORITY APPLICATION DATA:  
CC APPLICATION NUMBER: 07/353,432  
CC FILING DATE: 18-MAY-1989  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Goldstein, Jorge A.  
CC REGISTRATION NUMBER: 29,021  
CC TELEPHONE: (202) 371-2600  
CC TELEFAX: (202) 371-2540  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 2608 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: both  
CC TOPOLOGY: both  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 15..2363  
CC  
CC SEQUENCE 2608 BP: 557 A; 750 C; 804 G; 497 T; 0 OTHER.  
SQ  
Query Match 5.2%; Score 138; DB 1; Length 2608;  
Best Local Similarity 36.8%; Pred. No. 1.00e+00;  
Matches 32; Conservative 19; Mismatches 36; Indels 0; Gaps 0;

DB 2060 AGAGTTCCTGCTGATGACATGACAGAGACCTGGCAACGATGACGATCTTCTGGA 2119  
CP 412 WNAWSWYTCNGCRAAYTCRTTNARRANGCYTGNGTCKNSWRANGTGTGNGGTGTA 353  
DB 2120 CACCTGGACCAAGCTTGTCTGGGT 2146  
CP 352 ANACNGCNAACNACCTCTTGTGNGGRT 326

RESULT 8  
ID US-08-413-118-118 STANDARD; DNA; UNC; 1760 BP.  
AC xxxxxx  
DT  
DE Sequence 118, Application US/08413118  
CC Sequence 118, Application US/08413118  
CC Patent No. 5688920  
CC GENERAL INFORMATION:  
CC APPLICANT: PROLETTI, ENZO  
CC APPLICANT: LIMBACH, KEITH J.  
CC TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
CC CANINE HERPESVIRUS 9B, 9C, AND 9D AND USES THEREFOR  
CC NUMBER OF SEQUENCES: 128  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.  
CC STREET: 530 FIFTH AVENUE, 25TH FLOOR  
CC CITY: NEW YORK  
CC STATE: NEW YORK  
CC COUNTRY: UNITED STATES OF AMERICA  
CC ZIP: 10036  
CC  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk



```

CC APPLICATION NUMBER: US/08/426,036
CC FILING DATE: 21-APR-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Mcgarigle Jr., Philip L.
CC REGISTRATION NUMBER: 31,395
CC REFERENCE/DOCKET NUMBER: 0681.014
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (510) 601-2718
CC TELEFAX: (510) 655-3542
CC TELEX: n/a
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1642 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA (genomic)
CC FEATURE:
CC NAME/KEY: CDS
CC LOCATION: 179..946
CC FEATURE:
CC NAME/KEY: mat_peptide
CC LOCATION: 275..946
CC FEATURE:
CC NAME/KEY: -
CC LOCATION: 340
CC OTHER INFORMATION: /note="Intron Sequence"
SQ SEQUENCE 1642 BP; 356 A; 480 C; 403 G; 403 T; 0 OTHER.

Query Match 5.0%; Score 132; DB 1; Length 1642;
Best Local Similarity 34.0%; Pred. No. 2,82e+00;
Matches 32; Conservative 25; Mismatches 37; Indels 0; Gaps 0.

Db 1053 ACCAGCTCAGAGAGACAGACAGTGGAGCTGTACCTCTTGATATGACAGATATCTTCT 1112
Oy 483 HCAGATATTNATHATYARATHGARGGNCWMTYNTATATATGARGATYMSNATHATYGT 542
Db 1113 ATTGTGCAGATTAAAGATTGCATTAGCTTTTTC 1146
Oy 543 NYTGARGCARTTTCATYAAVAGCGNGTNTNTTY 576

RESULT 10
ID US-08-401-013-3 STANDARD; DNA; UNC; 1642 BP.
AC xxxxxx

DE Sequence 3, Application US/08401013
CC Sequence 3, Application US/08401013
CC Patent No. 5681719
CC GENERAL INFORMATION:
CC APPLICANT: LADNER, MARTHA B.
CC APPLICANT: NOBLE, JANELLE A.
CC APPLICANT: MARTIN, GEORGE A.
CC APPLICANT: KAWASAKI, ERNEST S.
CC APPLICANT: COYNE, MAZIE YEE
CC APPLICANT: HALENBECK, ROBERT F.
CC APPLICANT: KOHS, KIRSTON E.
CC TITLE OF INVENTION: NEW FORMS OF COLONY STIMULATING FACTOR-1
CC NUMBER OF SEQUENCES: 24
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: CHIRON CORPORATION
CC STREET: Intellectual Property - R440, P.O. Box 8097
CC CITY: Emeryville
CC STATE: California
CC COUNTRY: U.S.A.
CC ZIP: 94662-8097
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:

```



```
CC APPLICATION NUMBER: US/08/4401,013
CC FILING DATE: 08-MAR-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/999,280
CC FILING DATE: 28-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: McGarrigle Jr., Phillip L.
CC REGISTRATION NUMBER: 31,395
CC REFERENCE/DOCKET NUMBER: 0681.007
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (510) 601-2718
CC TELEFAX: (510) 655-3542
CC TELEX: N/A
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1642 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA (genomic)
CC FEATURE:
CC NAME/KEY: CDS
CC LOCATION: 179..946
CC FEATURE:
CC NAME/KEY: mat_peptide
CC LOCATION: 275..946
CC FEATURE:
CC NAME/KEY: -
CC LOCATION: 340
CC OTHER INFORMATION: /note= "Intron Sequence"
S0 SEQUENCE 1642 BP; 356 A; 480 C; 403 G; 403 T; 0 OTHER.

Query Match          5.0%; Score 132; DB 2; Length 1642;
Best Local Similarity 34.0%; Pred. No. 2.82e+00;
Matches    32; Conservative   25; Mismatches 37; Indels   0; Gaps   0;

Db 1053 ACCAGCTCAGAGAGACGACGTGGGACTGTACCTTCCTTGATATGCAGCATTTCTTCT 1112
QY 483 HCAGGAAYTNATHCAVYAARATHGARCGNGCMWSNMTNATHAAYGARGAYSNATHAAVGT 542
DY 1113 ATTTGTCAGATTAAAGTTCATTAGTTTTTTC 1146
QY 543 NYTNGARCARTTYGAYNAAYCGNGTNGTNTNTITY 576

RESULT      11
AC ID US-08-354-456A-3 STANDARD; DNA; UNC; 1642 BP.
AC xxxxxx

DE Sequence 3, Application US/08354456A
CC Sequence 3, Application US/08354456A
CC Patent No. 5567611
CC GENERAL INFORMATION:
CC APPLICANT: Ralph, Peter
CC APPLICANT: Martin, George
CC APPLICANT: Platek, Michael
CC APPLICANT: Larrick, James W.
CC TITLE OF INVENTION: Multifunctional M-CSF Proteins and Genes Encoding
CC TITLE OF INVENTION: Therefor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: CHIRON CORPORATION
CC STREET: Intellectual Property - R440, P.O. Box 8097
CC CITY: Emeryville
CC STATE: California
CC COUNTRY: U.S.A.
CC ZIP: 94662-8097
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
```

```

CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/354,456A
CC      FILING DATE: 12-DEC-1994
CC      CLASSIFICATION: 435
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 07/995,338
CC      FILING DATE: 21-DEC-1992
CC      CLASSIFICATION: 435
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: McGarrigle Jr., Philip L.
CC      REGISTRATION NUMBER: 31,395
CC      REFERENCE/DOCKET NUMBER: 750,003/32387
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (510) 601-2718
CC      TELEFAX: (510) 655-3542
CC      TELEX: n/a
CC      INFORMATION FOR SEQ ID NO: 3:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 1642 base pairs
CC      TYPE: nucleic acid
CC      STRANDEDNESS: single
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: DNA (genomic)
CC      FEATURE:
CC      NAME/KEY: CDS
CC      LOCATION: 179..946
CC      FEATURE:
CC      NAME/KEY: mat_peptide
CC      LOCATION: 275..946
CC      FEATURE:
CC      NAME/KEY:
CC      LOCATION: 340
CC      OTHER INFORMATION: /note="Intron Sequence"
SQ      SEQUENCE 1642 BP; 356 A; 480 C; 403 G; 403 T; 0 OTHER.

Query Match          5.0%; Score 132; DB 1; Length 1642;
Best Local Similarity 34.0%; Pred. No. 2.82e+00;
Matches 37; Conservative 25; Mismatches 37; Indels 0; Gaps 0.

Db 1053 ACCAGCTCAGAGAGAGTACAGTGGAGGAGTACCTCTTGATATGACAGTATCTTCT 1112
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 483 HCARAYVYTNATHGAYARATHAGARGNGCWMVYTNATHAYGARGAYSNATHAYGT 542
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1113 ATTGTGCGAGATTAAGATTGCATTAGTTTTC 1146
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 543 NYTNARGCARFTYGAAYAGCGNGTNGTNYTWTY 576

RESULT 12
ID US-07-999-280A-3 STANDARD: DNA; UNC; 1642 BP.
DT
AC xxxxxx
DE Sequence 3, Application US/07999280A
CC Sequence 3, Application US/07999280A
CC Patent No. 5573930
CC GENERAL INFORMATION:
CC APPLICANT: LADNER, MARTHA B.
CC APPLICANT: NOBLE, JANELLE A.
CC APPLICANT: MARTIN, GEORGE A.
CC APPLICANT: KAMASAKI, ERNEST S.
CC APPLICANT: COYNE, MAZIE YEE
CC APPLICANT: HALENBECK, ROBERT F.
CC APPLICANT: KOTHS, KIRSTON E.
CC TITLE OF INVENTION: NEW FORMS OF COLONY STIMULATING FACTOR-1
CC NUMBER OF SEQUENCES: 24
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: CHIRON CORPORATION
CC STREET: Intellectual Property - R440, P.O. Box 8097
CC CITY: Emeryville
CC STATE: California
CC COUNTRY: U.S.A.
CC ZIP: 94662-8097
CC COMPUTER READABLE FORM:

```



CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/999,280A  
CC FILING DATE: 28-DEC-1992  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: McGarrigle Jr., Phillip L.  
CC REGISTRATION NUMBER: 31,395  
CC REFERENCE/DOCKET NUMBER: 0681.007  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (510) 601-2718  
CC TELEFAX: (510) 655-3542  
CC TELEX: n/a  
CC INFORMATION FOR SEQ ID NO: 3:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1642 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 179..946  
CC FEATURE:  
CC NAME/KEY: mat-peptide  
CC LOCATION: 275..946  
CC FEATURE:  
CC NAME/KEY:  
CC LOCATION: 340  
CC OTHER INFORMATION: /note="Intron Sequence"  
CC SEQUENCE 1642 BP; 356 A; 480 C; 403 G; 403 T; 0 OTHER.

Query Match 5.0%; Score 132; DB 1; Length 1642;  
Best Local Similarity 34.0%; Pred. No. 2.82e+00;  
Matches 32; Conservative 25; Mismatches 37; Indels 0; Gaps 0;

Db 1053 ACCAGCTCAGAGAGATGACGCTGTTACCTTCCTTGATATGACAGTATTTCTT 1112  
Qy 483 HCARGAVYTNATGAAATGATGARGGNCMSNTNATHAAGARGAVSMNTHAAYGT 542  
Db 1113 ATTGTGACAGTAAAGTTCATTAGTTTTC 1146  
Qy 543 NTNGARCARTTGAATGATGCGTNGTNGTNTT 576

RESULT 13  
ID US-08-426-279-3 STANDARD; DNA; UNC; 1642 BP.  
AC xxxxxx  
DE Sequence 3, Application US/08426279  
CC Sequence 3, Application US/08426279  
CC Patent No. 5672343  
CC GENERAL INFORMATION:  
CC APPLICANT: LADNER, MARTHA B.  
CC APPLICANT: NOBLE, JANELLE A.  
CC APPLICANT: MARTIN, GEORGE A.  
CC APPLICANT: KAWASAKI, ERNEST S.  
CC APPLICANT: COYNE, MAZIE YEE  
CC APPLICANT: HALENEBECK, ROBERT F.  
CC APPLICANT: KOHNS, KIRSTON E.  
CC TITLE OF INVENTION: NEW FORMS OF COLONY STIMULATING FACTOR-1  
CC NUMBER OF SEQUENCES: 24  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: CHIRON CORPORATION  
CC STREET: Intellectual Property - R440, P.O. Box 8097  
CC CITY: Emeryville  
CC STATE: California  
CC COUNTRY: U.S.A.  
CC ZIP: 94662-8097  
CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/426,279  
CC FILING DATE: 21-APR-1995  
CC CLASSIFICATION: 530  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: McGarrigle Jr., Phillip L.  
CC REGISTRATION NUMBER: 31,395  
CC REFERENCE/DOCKET NUMBER: 0681.013  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (510) 601-2718  
CC TELEFAX: (510) 655-3542  
CC TELEX: n/a  
CC INFORMATION FOR SEQ ID NO: 3:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1642 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 179..946  
CC FEATURE:  
CC NAME/KEY: mat-peptide  
CC LOCATION: 275..946  
CC FEATURE:  
CC NAME/KEY:  
CC LOCATION: 340  
CC OTHER INFORMATION: /note="Intron Sequence"  
CC SEQUENCE 1642 BP; 356 A; 480 C; 403 G; 403 T; 0 OTHER.

Query Match 5.0%; Score 132; DB 2; Length 1642;  
Best Local Similarity 34.0%; Pred. No. 2.82e+00;  
Matches 32; Conservative 25; Mismatches 37; Indels 0; Gaps 0;

Db 1053 ACCAGCTCAGAGAGATGACGCTGTTACCTTCCTTGATATGACAGTATTTCTT 1112  
Qy 483 HCARGAVYTNATGAAATGATGARGGNCMSNTNATHAAGARGAVSMNTHAAYGT 542  
Db 1113 ATTGTGACAGTAAAGTTCATTAGTTTTC 1146  
Qy 543 NTNGARCARTTGAATGATGCGTNGTNGTNTT 576

RESULT 14  
ID US-08-231-729B-2 STANDARD; DNA; UNC; 1659 BP.  
AC xxxxxx  
DE Sequence 2, Application US/08231729B  
CC Sequence 2, Application US/08231729B  
CC Patent No. 5618722  
CC GENERAL INFORMATION:  
CC APPLICANT: ZENNO, Shuhei  
CC APPLICANT: SHIRAIISHI, Shinji  
CC APPLICANT: INOUE, Satoshi  
CC APPLICANT: SAIGO, Kaoru  
CC TITLE OF INVENTION: FIREFLY LUCIFERASE GENE  
CC NUMBER OF SEQUENCES: 8  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: LEYDIG, VOIT & MAYER  
CC STREET: 700 Thirteenth Street, N.W., Suite 300  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20005  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS



CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/231,729B  
CC FILING DATE: 20-APR-1994  
CC CLASSIFICATION: 435  
CC PRIORITY APPLICATION DATA:  
CC APPLICATION NUMBER: JP 119050/1993  
CC FILING DATE: 21-APR-1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Rose, Herbert C.  
CC REGISTRATION NUMBER: 29846  
CC REFERENCE/DOCKET NUMBER: 60130/No. 5618722aka  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 202-737-6770  
CC TELEFAX: 202-737-6776  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1659 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: CDNA  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 1..1659  
CC OTHER INFORMATION: for codons 28, 32, 112, 130, 142, 190, 212,  
CC OTHER INFORMATION: 217, 222, 266, 329, 336, 386, 436, 512, and 532, if the 3'  
CC OTHER INFORMATION: nucleotide is T or C, then the 5' nucleotide is C; and if  
CC OTHER INFORMATION: nucleotide is A, then the 3' nucleotide is A or G  
SQ SEQUENCE 1659 BP: 321 A; 193 C; 268 G; 265 T; 612 OTHER.

Query Match 5.0%; Score 132; DB 1; Length 1659;  
Best Local Similarity 20.0%; Pred. No. 2.82e+00;  
Matches 16; Conservative 30; Mismatches 34; Indels 0; Gaps 0;

Db 1576 ACONGNARATHGAYACNMGNAARATHAARGARATHYTNATHAARGCNCARARGNAR 1635  
OY 8 AYWSGAYGAYATHAYGCNCARATHHTTYCARATHACNGAYAAARGNACNGCNGTNGAYG 67  
DB 1636 TCNARTCNARGCNARTT 1655  
OY 68 TTTAYGTNGAYGNGARTTY 87

RESULT 15  
ID US-08-231-729B-1 STANDARD; DNA; UNC; 1659 BP.  
AC xxxxxx  
DT Sequence 1, Application US/08231729B  
DE Sequence 1, Application US/08231729B  
CC Patent No. 5618722  
CC GENERAL INFORMATION:  
CC APPLICANT: ZENNO, Shuhel  
CC APPLICANT: SHIRAIISHI, Shinji  
CC APPLICANT: INOUE, Satoshi  
CC APPLICANT: SAIGO, Kaoru  
CC TITLE OF INVENTION: FIREFLY LUCIFERASE GENE  
CC NUMBER OF SEQUENCES: 8  
CC CORRESPONDENCE ADDRESSES:  
CC ADDRESSEE: LEYDIG, VOIT & MAYER  
CC STREET: 700 Thirteenth street, N.W., Suite 300  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20005  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/231,729B  
CC FILING DATE: 20-APR-1994

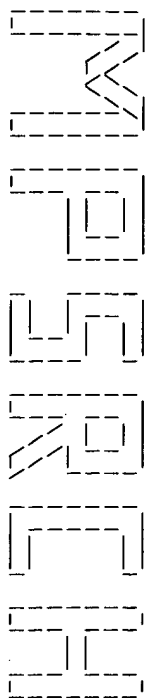
CC CLASSIFICATION: 435  
CC PRIORITY APPLICATION DATA:  
CC APPLICATION NUMBER: JP 119050/1993  
CC FILING DATE: 21-APR-1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Rose, Herbert C.  
CC REGISTRATION NUMBER: 29846  
CC REFERENCE/DOCKET NUMBER: 60130/No. 5618722aka  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 202-737-6770  
CC TELEFAX: 202-737-6776  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1659 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: CDNA  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 1..1659  
CC OTHER INFORMATION: "Xaa" at codon 409 is either Glu or Asp;  
CC OTHER INFORMATION: "Xaa" at codons 28, 32, 112, 130, 142, 190, 212, 217, 2  
CC OTHER INFORMATION: 329, 336, 386, 436, 512, and 532 is either Arg, Ser or  
SQ SEQUENCE 1659 BP: 321 A; 193 C; 268 G; 265 T; 612 OTHER.

Query Match 5.0%; Score 132; DB 1; Length 1659;  
Best Local Similarity 20.0%; Pred. No. 2.82e+00;  
Matches 16; Conservative 30; Mismatches 34; Indels 0; Gaps 0;

Db 1576 ACONGNARATHGAYACNMGNAARATHAARGARATHYTNATHAARGCNCARARGNAR 1635  
OY 8 AYWSGAYGAYATHAYGCNCARATHHTTYCARATHACNGAYAAARGNACNGCNGTNGAYG 67  
DB 1636 TCNARTCNARGCNARTT 1655  
OY 68 TTTAYGTNGAYGNGARTTY 87

Search completed: Sat Nov 27 10:19:43 1999  
Job time : 53 secs.





(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Nov 24 00:48:01 1999; Maspar time 11.00 Seconds  
Tabular output not generated. 783.051 Million cell updates/sec

Title: >US-09-103-287-4  
Description: (1-215) from US09103287.pep  
Perfect Score: 1495  
Sequence: 1 FKSDSDIYAGIFQITDKGTA.....GDIOKLNAYIDKGMKNAP 215

Scoring table: PAM 150  
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 47.121; Variance 120.640; scale 0.391

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	863	57.7	432	2	C69662	UDP-N-acetylmuramate-	4.56e-113
2	366	23.8	495	2	C71679	UDP-N-acetylmuramate-	1.01e-33
3	343	22.9	505	2	S76722	UDP-N-acetylmuramate-	8.51e-32
4	341	22.8	454	2	B70418	UDP-N-acetylmuramate-	1.68e-31
5	293	19.6	803	2	A71475	Probable muramate-Ala	1.70e-24
6	291	18.9	475	2	E64185	UDP-N-acetylmuramate-	3.31e-24
7	283	18.5	468	2	H70201	UDP-N-acetylmuramate-	4.65e-23
8	264	17.7	481	2	C71338	Probable UDP-N-acetyl	2.35e-20
9	259	17.3	444	2	D70579	Probable murc protein	1.19e-19
10	239	16.0	449	2	B71917	UDP-N-acetylmuramate-	7.40e-17
11	235	15.7	449	2	G64597	UDP-N-acetylmuramate-	2.64e-16
12	232	15.5	491	1	CEECAM	UDP-N-acetylmuramate-	6.83e-16
13	193	12.9	457	2	S56459	hypothetical 48.5k pr	1.20e-10
14	185	12.4	453	2	B64002	hypothetical protein	1.32e-09
15	150	10.0	682	2	B69170	UDP-N-acetylmuramyl t	3.20e-05
16	132	8.8	494	2	B47691	UDP-N-acetylmuramoyla	4.14e-03
17	130	8.7	505	2	S75968	UDP-N-acetylmuramoyla	6.98e-03
18	128	8.6	450	2	B69198	UDP-N-acetylmuramyl t	1.17e-02
19	127	8.5	1558	2	B71603	REBA-H3 antigen PRB09	1.52e-02
20	118	7.9	358	2	D71127	hypothetical protein	1.48e-01
21	114	7.6	457	2	I64184	UDP-N-acetylmuramoyla	3.94e-01
22	109	7.3	143	2	E69065	molybdenum cofactor b	1.31e+00
23	104	7.0	283	2	A71927	hypothetical protein	4.19e+00

24	104	7.0	447	2	D71812	udp-n-acetylmuramyl-t	4.19e+00
25	103	6.9	447	2	F64706	UDP-MurNac-tripeptide	5.27e+00
26	103	6.9	881	2	S37895	LH51 protein precursor	5.27e+00
27	102	6.8	452	2	UC6561	UDP-N-acetylmuramoyla	6.61e+00
28	102	6.8	477	2	A49937	50k outer membrane pr	6.61e+00
29	101	6.8	591	2	S70524	guanine nucleotide-bi	8.29e+00
30	102	6.8	948	2	JC2190	preprotein translocas	6.61e+00
31	100	6.8	1092	1	S37676	glutamate dehydrogena	8.29e+00
32	100	6.7	420	2	S75514	sensory transduction	1.04e+01
33	100	6.7	532	2	D71267	probable UDP-N-acetyl	1.04e+01
34	100	6.7	539	2	S67049	probable membrane pro	1.04e+01
35	100	6.7	585	2	C70330	conserved hypothetical	1.04e+01
36	100	6.7	591	2	S43506	hypothetical protein	1.04e+01
37	100	6.7	1013	2	J01920	DNA-directed RNA poly	1.04e+01
38	98	6.6	185	1	RNV222	DNA-directed RNA poly	1.62e+01
39	99	6.6	445	2	G70371	UDP-MURNAC-pentapepti	1.30e+01
40	99	6.6	508	2	A70125	UDP-N-acetylmuramoyla	1.30e+01
41	98	6.6	589	2	A46459	macrophage-activation	1.62e+01
42	99	6.6	1235	2	C71210	probable DNA-directed	1.30e+01
43	97	6.5	185	2	F36845	DNA-directed RNA poly	2.01e+01
44	97	6.5	276	2	G64584	conserved hypothetical	2.01e+01
45	97	6.5	914	2	JC3574	inter-alpha-trypsin I	2.01e+01

## ALIGNMENTS

RESULT	1
ENTRY	C69662
TITLE	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murc - Bacillus subtilis
ORGANISM	#formal name Bacillus subtilis
DATE	05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 17-Mar-1999
ACCESSIONS	C69662; S71002
REFERENCE	A69580
#authors	Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Biollet, A.; Borchert, S.; Bross, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capriano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conterton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoef, A.; Ehrlich, S.D.; Emmerson, P.T.; Ertlin, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleron, N.; Ghm, S.Y.; Glaeser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaber-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Mostl, D.; Nakai, S.; Noback, M.; Moore, D.; O'Reilly, M.; Ogawa, K.; Ogihara, A.; Oudesa, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetlelle, D.; Porwolik, S.; Prescott, A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serro, P.; Shin, B.S.; Solido, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takenuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambut, R.; Wedler, E.; Wedler, H.; Weitzneger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

#journal  
#title  
The complete genome sequence of the Gram-positive bacterium  
Bacillus subtilis.



[illegible]



```

OY 98 ISATIDTARKKKYPH-K-E-VVAVPQPHHFSRTQAFLNFEASLSKADRVFLCEIFGSIRES 154
Db 442 NPYNIRGEDL 451
OY 155 NTGALTIDOL 164

RESULT 4
ENTRY B70418 #type complete
TITLE UDP-N-acetylmutamate-alanine ligase - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change 21-Aug-1998

ACCESSIONS B70418
REFERENCE B70300
#authors Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.O.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
#cross-references MUID:98196666
#accession B70418
#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type DNA
#residues 1-454 #label AOF
#cross-references GB:AE000736; NID:g2983763; PID:g2983764; GB:AE000657
#experimental_source strain VFS

GENETICS
#gene mure
CLASSIFICATION #superfamily UDP-N-acetylmutamate--alanine ligase
SUMMARY #length 454 #molecular_weight 50893 #checksum 2346

Query Match 22.8%; Score 341; DB 2; Length 454;
Best Local Similarity 34.1%; Pred. No. 1,08e-31;
Matches 61; Conservative 48; Mismatches 62; Indels 8; Gaps 8;

Db 276 GIHNVYNNALATGVA-LE-LGVSFEVKKSLSEEFNARERLELGGYKNSPVDDYGHHP 333
OY 39 GDHVLNMLAVAIATYLEKLDVT-N-IKEALETFGGVARRNETTI-ANQYIVDDYVHHHP 95
Db 334 TEIKAVINSLRDMPDKNLVLVFPDRYSRTYYLFEDEFVKYLDKIDKLIYDIYASENN 393
OY 96 REISATIDTARKKKYPHKEVAVFQGHFHSRQAFLNFEASLSKADRVFLCEIFGSIRES 155
Db 394 VYGVSABELARKSGAVFPAKDKKEVFEEKYREVHDGSDVILTFAGSISISKWCEEFLEKYNL 452
OY 156 TGAALTIDOLIDIKIEGA-SLINEDSIN-VLLEOFDNA-VVLEFAGADIDKLONAVYLDKLGIM 211

RESULT 5
ENTRY A71475 #type complete
TITLE Probable muramate-Ala ligase and D-Ala-D-Ala ligase - Chlamydia trachomatis (serotype D, strain uw3/Cx)
ORGANISM #formal_name Chlamydia trachomatis
DATE 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 21-Nov-1998

ACCESSIONS A71475
REFERENCE A71570
#authors Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov, R.L.; Zhao, O.; Koonin, E.V.; Davis, R.W.
#journal Science (1998) 282:754-759
#title Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.
#cross-references MUID:9900809
#accession A71475
#status preliminary
#molecule_type DNA
#residues 1-803 #label ARN
#cross-references GB:AE001348; GB:AE001273; NID:g3329216; PID:g3329222

```

```

GENETICS ##experimental_source serotype D, strain UW-3/Cx
#gene muc/ddIA
SUMMARY #length 803 #molecular_weight 89229 #checksum 235

Query Match 19.6%; Score 293; DB 2; Length 803;
Best Local Similarity 32.6%; Pred. No. 1.70e-24;
Matches 45; Conservative 42; Mismatches 49; Indels 2; Gaps 2;

Db 266 GKHNVLNAAAMGIALSGIDEGAIRNAFRGFSGVQRILORKNKSSEFFLEDDYAHHPSE 325
| | |||| : : : : : | : : : : : | : : : : : | : : : : : |
Qy 39 GDHYVLNLAVIATLSYLEKLDVTNIKEALEFFGSVKRRFMTITANO-VIVDDYAHHPRE 97
| | : : : : : | : : : : : | : : : : : | : : : : : |

Db 336 ISCTLRAVTVAGVORRIATYOPIHRSFLRRCIDSPSFAEKDADEVILTVEYSAG-EAE 364
| | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 98 ISATIDTARKKYPRHREVAVPQPHTSFRTQAFLNEFASLTKADRVFLCEIFGSIRENTG 157
| | : : : : : | : : : : : | : : : : : | : : : : : |

Db 365 DISYOKLAELAISQESIYK 402
| | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 158 ALTIODLLDKIEGASLIN 175
| | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 6
ENTRY E64185 #type complete
TITLE UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) -
ORGANISM Haemophilus influenzae (strain Rd kw20)
#formal_name Haemophilus influenzae
DATE 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
24-Oct-1998

ACCESSIONS E64185
REFERENCE A64000
#authors Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
Kirchess, E.F.; Kirlavage, A.R.; Bult, C.J.; Tomb, J.F.;
Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
Flitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weisman,
J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann
J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
Science (1995) 269:496-512
Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd.

#cross-references MUID:95350630
#accesion E64185
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-475 #label TIGR
##cross-references GB:U3794; GB:I42023; NID:g1574694; PID:g1574695;
TIGR:H1139

FUNCTION
#description one of the ligases responsible for the synthesis of
UDP-N-acetylmuramyl pentapeptide, an intermediate in
cell-wall biosynthesis
peptidoglycan biosynthesis
CLASSIFICATION #superfamily UDP-N-acetylmuramate--alanine ligase
KEYWORDS ATP; cell division; cell wall; ligase; P-loop; peptidoglycan
biosynthesis

FEATURE
125-130 #region nucleotide-binding motif A (P-loop)
SUMMARY #length 475 #molecular_weight 51994 #checksum 7812

Query Match 19.5%; Score 291; DB 2; Length 475;
Best Local Similarity 34.6%; Pred. No. 3.31e-24;
Matches 62; Conservative 48; Mismatches 51; Indels 18; Gaps 15;

Db 289 GKHNMLNATAALAAKEGIANEAILEALDFOGAGRPFOLGEFIRPNCKVRLVDYGH 348
| | |||| : : : : : | | | | : : : : : | : : | : : ||||: |
Qy 39 GDHYVLNLAVIATLSYLEKLDVTNIKEALEFFGSVKRRFNF-IT-IAN-QV-IYDDYAH 93
| | : : : : : | : : : : : | : : : : : | : : : : : |

Db 349 HPTTEGVTTIKARREGWGDKRIVMTFOPHRRSRTDLDEFDVQLSDVALIMLVYAAGE 408
| | : : : : : | : : : : : | : : : : : | : : : : : |

```



```

OY 94 HPREISATIDARKKYPKEVAVFQPHFTSRTQAFLEAFESLSKADRVLCETFF--G- 150
Db 409 APIVAGADSKSLORSIRNL-GRVD-PIIVSDTSQGLDVLDDIOTDGDLLIAGAGSVSKI 465
OY 151 S-IR-ENTGAL--TIQDLIDIKEGASLINEDS-I-NVLEQF--DNAVYLFVAGAGDIQKL 201

RESULT 7
ENTRY 7
TITLE H70201 #type complete
ORGANISM UDP-N-acetylmuramate--alanine ligase (murC) homolog - Lyme
#formal_name Borrelia burgdorferi #common_name Lyme disease
#spirochete
DATE 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
05-Jun-1998
ACCESSIONS H70201
REFERENCE A70100
#authors Fraser, C.M.; Castens, S.; Huang, W.M.; Sutton, G.G.;
Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
Dodson, R.; Hickey, E.K.; Gwin, M.; Dougherty, B.; Tomb,
J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
Vogt, R.Y.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
J.; Utterback, T.; Wathey, L.; McDonald, L.; Artach, P.;
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
#journal Nature (1997) 390:580-586
#title Genomic sequence of a Lyme disease spirochete, Borrelia
burgdorferi.
#cross-references MUID:98065943
#accession H70201
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-468 #label KLE
##cross-references GB:AE001180; GB:AE000783; NID:g2688755; PID:g2688761;
TIGR:DB0817
##experimental_source strain B31
SUMMARY #length 468 #molecular_weight 53480 #checksum 8199
Query Match 18.9%; Score 283; DB 2; Length 468;
Best Local Similarity 38.3%; Pred. No. 4,65e-23;
Matches 59; Conservative 33; Mismatches 51; Indels 11; Gaps 8;

Db 316 IKRIAKNSGIRREVEVYKGYIMDYAHHPREIKNTLFGIKNFYKRNRIIDEMPH 375
OY 63 IKELLEFGGKRRFNETTIANQYI-VDDVAHHPREISATIDTARKKYPKEVAVFQPH 121
Db 376 TFFTKKEFFADFEVLNADILIHNIYLSNRENFPELSVKLPLIKTKNTYFFKD 435
OY 122 TFSRTQAFLEAFESLSKADRVLCETFFGSIRES--TGALTID-L-IDKIEGAS-LINE 176
Db 436 VKDSINFIKSLISGDLFITMGAGNNFILD-FL 468
OY 177 --DSINVLQDFDNVLF--WGAGDIQKLQNAVYL 206

RESULT 8
ENTRY 8
TITLE C71338 #type complete
ORGANISM Probable UDP-N-acetylmuramate--alanine ligase (murC) -
#formal_name Treponema pallidum subsp. pallidum #common_name
syphilis spirochete
DATE 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
17-Mar-1999
ACCESSIONS C71338
REFERENCE A71350
#authors Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.;
Sutton, G.G.; Dodson, R.; Gwin, M.; Hickey, E.K.; Clayton,
R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod,
M.P.; Salzberg, S.; Peterson, J.; Khalak, H.; Richardson,
D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald,
L.; Artach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.;

```

```

Garland, S.; Hatch, B.; Horst, K.; Roberts, K.; Wathey,
J.; Weidman, J.; Smith, H.O.; Venter, J.C.
#journal Science (1998) 281:375-388
#title Complete genome sequence of Treponema pallidum, the syphilis
spirochete
#cross-references MUID:98332770
#accession C71338
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-481 #label COL
##cross-references GB:AE001213; GB:AE000520; NID:g3322606; PID:g3322616
##experimental_source strain Nichols
GENETICS TP0341
SUMMARY #length 481 #molecular_weight 53762 #checksum 6883
Query Match 17.7%; Score 264; DB 2; Length 481;
Best Local Similarity 34.0%; Pred. No. 2.35e-20;
Matches 50; Conservative 35; Mismatches 55; Indels 7; Gaps 7;

Db 280 GEFVATRGESVLANATGALALSLVKKOYGEVTVHATKRVYALFGCCRRSEVLG 339
OY 27 GEFYDHLSPQYG-DHT-VLN-ALAVTAISLE-KLD-VINIKALETFEGVKKREVE-T 80
Db 340 EVRGILFMDYDGHHPATKRLKLTFFERRIVDFMSHTSRTAALTERRAESFQDA 399
OY 81 TIANQVYDDVAHHPREISATIDTARKKYPKEVAVFQPHFTSRTQAFLEAFESLSKA 140
Db 400 DVTILHEITVASAREVYQGEVNGEHLFE 426
OY 141 DRVFLCEIFGSIRE-NTGALTIDLDLD 166

RESULT 9
ENTRY 9
TITLE D70579 #type complete
ORGANISM Probable murC protein - Mycobacterium tuberculosis (strain
H37RV)
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS D70579
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry,
III, C.E.; Tekait, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornby, T.; Jagsels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squares, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession D70579
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-494 #label COL
##cross-references GB:295388; GB:AL123456; NID:g3261759; PID:e315954;
PID:g2104326
##experimental_source strain H37RV
GENETICS murC
SUMMARY #length 494 #molecular_weight 51176 #checksum 1162
Query Match 17.3%; Score 259; DB 2; Length 494;
Best Local Similarity 31.9%; Pred. No. 1.19e-19;
Matches 60; Conservative 44; Mismatches 69; Indels 15; Gaps 11;

```







uridine-diphosphate-N-acetylmuramate-L-alanine ligase from  
Escherichia coli.  
#cross-references MUID:95324553  
#accession S65354  
##status preliminary  
##molecule-type protein  
##residues 1-14 #label LIG

REFERENCE  
#authors A64720  
Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;  
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;  
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;  
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,  
Y.

#journal Science (1997) 277:1453-1462  
#title The complete genome sequence of Escherichia coli K-12.  
#cross-references MUID:97426617  
#accession C64731  
##status preliminary; nucleic acid sequence not shown;  
translation not shown

##molecule-type DNA  
##residues 1-491 #label BLAT  
##cross-references GB:AE000118; GB:U00096; NID:g1786262; PID:g1786279;  
UMGP:b0091

##experimental\_source strain K-12, substrain MG1655

GENETICS  
#gene murC  
#map\_position 2 min

FUNCTION  
#description one of the ligases responsible for the synthesis of  
UDP-N-acetylmuramyl pentapeptide, an intermediate in  
cell-wall biosynthesis  
peptidoglycan biosynthesis  
#pathway  
#superfamily UDP-N-acetylmuramate--alanine ligase  
ATP; cell division; cell wall; ligase; P-loop; peptidoglycan  
biosynthesis

FEATURE  
126-131 #region nucleotide-binding motif A (P-loop)  
SUMMARY #length 491 #molecular-weight 53626 #checksum 8296

Query Match 15.5%; Score 232; DB 1; Length 491;  
Best Local Similarity 38.8%; Pred. No. 6.83e-16;  
Matches 47; Conservative 28; Mismatches 36; Indels 10; Gaps 5;

Db 290 GHHNLAANAATAATGECIDDEALIRALIESQGTGRPFDEGPELPVNGKSGTAMLV 349  
| | | | | : : : : : | | | | | : : : : : | : : : : :  
Qy 39 GDHYTLNLAVALAISYLELDVTNIKEALETFGVKRRFN--E-T--TI--AN-QVIV 88  
| | | | | : : : : : | | | | | : : : : : | : : : : :  
Db 350 DDYGHPEVDATIKAAARGPDKNLVLMFQPHRTFRDLDVDFRANVLTQVDTLMEV 409  
| | | | | : : : : : | | | | | : : : : : | : : : : :  
Qy 89 DDYAHPREISATIDTARKKYPHKEVAVFQPHTESRTOAFINERASLSKADRVFLCEI 148  
| | | | | : : : : : | | | | | : : : : : | : : : : :  
Db 410 Y 410  
Qy 149 F 149

RESULT 13  
ENTRY S56459 #type complete  
TITLE hypothetical 48.5k protein (fbp-pmba intergenic region) -  
ALTERNATE\_NAMES Escherichia coli  
ORGANISM #journal\_name Escherichia coli  
DATE 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change  
14-Nov-1997

ACCESSIONS S56459; D65235  
REFERENCE S56314  
#authors Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.;  
Blattner, F.R.  
#journal Nucleic Acids Res. (1995) 23:2105-2119  
#title Analysis of the Escherichia coli genome VI: DNA sequence of  
the region from 92.8 through 100 minutes.  
#cross-references MUID:95334362  
#accession S56459

##status preliminary; nucleic acid sequence not shown;  
translation not shown

##molecule-type DNA  
##residues 1-457 #label BUR  
##cross-references EMBL:U14003; NID:g1263172; PID:g537075  
#note The nucleotide sequence was submitted to the EMBL Data  
Library, August 1994

REFERENCE  
#authors A64720  
Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;  
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;  
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;  
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,  
Y.

#journal Science (1997) 277:1453-1462  
#title The complete genome sequence of Escherichia coli K-12.  
#cross-references MUID:97426617  
#accession D65235  
##status preliminary; nucleic acid sequence not shown;  
translation not shown

##molecule-type DNA  
##residues 1-457 #label BLAT  
##cross-references GB:AE000494; GB:U00096; NID:g1790670; PID:g1790680;  
UMGP:b4233

##experimental\_source strain K-12, substrain MG1655

GENETICS  
#gene yjfg  
#map\_position 457

SUMMARY #length 457 #molecular-weight 49874 #checksum 3921

Query Match 12.9%; Score 193; DB 2; Length 457;  
Best Local Similarity 33.6%; Pred. No. 1.20e-10;  
Matches 37; Conservative 29; Mismatches 41; Indels 3; Gaps 3;

Db 277 GEHNNHNGILMAIAARHGVAPADANALGSGFINARRLELGEANGVTVYDDFAHPTA 336  
| | | | | : : : : : | | | | | : : : : : | : : : : :  
Qy 39 GDHYTLNLAVALAISYLELDVTNIKEALETFGVKRRFNFTTIANQYIV-DDYAHPRE 97  
| | | | | : : : : : | | | | | : : : : : | : : : : :  
Db 337 ILATLALRGVGVGTARIIVALEPRSNMTKMGICKDLAPSLGRADEVFL 386  
| | | | | : : : : : | | | | | : : : : : | : : : : :  
Qy 98 ISATIDTARKKYPHKE-VVAVFQPHTESRTOAF-LNEFAESLSKADRVFL 145  
| | | | | : : : : : | | | | | : : : : : | : : : : :  
RESULT 14  
ENTRY B64002 #type complete  
TITLE hypothetical protein H10121 - Haemophilus influenzae (strain  
Rd KW20)  
ORGANISM #journal\_name Haemophilus influenzae  
DATE 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change  
10-Oct-1997

ACCESSIONS B64002  
REFERENCE A64000  
#authors Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;  
Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.;  
Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;  
Fitzhugh, W.; Fields, C.; Gocayne, A.; Kelley, J.M.; Weidman,  
J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,  
M.D.; Usterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,  
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,  
J.L.; Geoghegan, N.S.M.; Gnehm, C.L.; McDonald, L.A.;  
Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.

#journal Science (1995) 269:496-512  
#title Whole-genome random sequencing and assembly of Haemophilus  
influenzae Rd.  
#cross-references MUID:95350630  
#accession B64002  
##status nucleic acid sequence not shown; translation not shown  
##molecule-type DNA  
##residues 1-453 #label TIGR  
##cross-references GB:U32698; GB:U42023; NID:g1573072; PID:g1573076;  
TIGR:HI0121

SUMMARY #length 453 #molecular-weight 50074 #checksum 6063

Query Match 12.4%; Score 185; DB 2; Length 453;



Best Local Similarity 29.4%; Pred. No. 1.32e-09;  
Matches 40; Conservative 33; Mismatches 60; Indels 3; Gaps 3;

[illegible]

RESULT	ENTRY	15
TITLE	A69170	#type complete
ORGANISM	UDP-N-acetylmuramyl tripeptide synthetase related protein -	
DATE	Methanobacterium thermoautotrophicum (strain Delta H)	
	#formal_name Methanobacterium thermoautotrophicum	
	05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change	
	05-Jun-1998	
ACCESSIONS	A69170	
REFERENCE	A69000	
#authors	Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.	

```

#molecule_type DNA
#residues 1-682 #label MTH
#cross-references GB:AE000836; GB:AE000666; NID:g2621601; PID:g2621604
#experimental_source strain Delta H

```

		10.0%;	Score 150;	DB 2;	Length 682;	
Query Match		Similarity	37.9%;	Pred.	No. 3,20e-05;	
Best Local	25;	Conservative	15;	Mismatches	25;	Indels 1; Gaps 1;
Matches						
DB	514	VYNALAAATAMTGMGLEIDIVRGLESEFKGVGRPQELSESPIRLD-YAHNPAGVRAYM	572			
		:	:::::	:	: : :       :	:
OY	43	VNALAAVAIAISYLEKLDVTINIKAELETGGVAKRRNETTIANOVIVDDYAAHPREISATI	102			
		:	:::::	:	: : :       :	:
DB	573	ODLRGK	578			
		: -				
OY	103	DTRKK	108			

Search completed: Wed Nov 24 00:49:00 1999  
Job time : 59 secs.



THIS PAGE INTENTIONALLY LEFT BLANK

**THIS PAGE BLANK (USPTO)**



(TM)

Distribution rights by Oxford Molecular Ltd

Tabular output not generated.

**Sequence:**

### Scoring table:

Searched:

### Statistics:

and is derived by analysis of the total score distribution

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1454	97.3	437	1	MURC_STA0	UDP-N-ACETYLUMRAMATE--	1.39e-24
2	863	57.7	432	1	MURC_BACSU	UDP-N-ACETYLUMRAMATE--	1.20e-13
3	343	22.9	505	1	MURC_SNY3	UDP-N-ACETYLUMRAMATE--	1.62e-37
4	337	21.9	433	1	MURC_PORG1	UDP-N-ACETYLUMRAMATE--	9.15e-35
5	303	20.3	280	1	MURC_SNP7	UDP-N-ACETYLUMRAMATE--	1.11e-30
6	291	19.5	475	1	MURC_HAEIN	UDP-N-ACETYLUMRAMATE--	1.17e-28
7	283	18.9	468	1	MURC_BORBU	UDP-N-ACETYLUMRAMATE--	2.56e-27
8	264	17.7	481	1	MURC_TREPA	UDP-N-ACETYLUMRAMATE--	3.62e-24
9	259	17.3	494	1	MURC_MTCU	UDP-N-ACETYLUMRAMATE--	2.40e-23
10	243	16.3	483	1	MURC_BOCAP	UDP-N-ACETYLUMRAMATE--	9.77e-21
11	235	15.7	449	1	MURC_HELPU	UDP-N-ACETYLUMRAMATE--	1.89e-19
12	232	15.5	491	1	MURC_ECOLI	UDP-N-ACETYLUMRAMATE--	5.72e-19
13	193	12.9	457	1	MPL_ECOLI	UDP-N-ACETYLUMRAMATE--L	7.16e-13
14	185	12.4	453	1	MPL_HAEIN	UDP-N-ACETYLUMRAMATE--L	1.76e-11
15	132	8.8	494	1	MURC_BACSU	UDP-N-ACETYLUMRAMATE--L	3.76e-04
16	130	8.7	505	1	MURC_SNY3	UDP-N-ACETYLUMRAMATE--L	6.87e-04
17	118	7.6	378	1	PGK_COMNG	PHOSPHOGLYCERATE KINAS	9.30e-02
18	114	7.6	457	1	MURC_HAEIN	UDP-N-ACETYLUMRAMATE--L	7.05e-02
19	104	7.0	417	1	PGK_CANAL	PHOSPHOGLYCERATE KINAS	1.05e+00
20	105	7.0	456	1	MURC_ENTPA	UDP-N-ACETYLUMRAMATE--L	8.09e-01
21	103	6.9	881	1	LHST_YEAST	HEAT SHOCK PROTEIN 70	1.37e+00
22	102	6.8	477	1	OPRK_PDEAE	OUTER MEMBRANE PROTEIN	1.77e+00
23	101	6.8	591	1	GBP2_HUMAN	INTERFERON-INDUCED GUA	2.29e+00

45	SECA_STMP7	PEPTOTENIN TRANSLOCASE	1.75e+00
24	DHE2_YEAST	NAD-SPECIFIC GLUTAMATE	2.29e+00
25	MCRD_TREPA	UDP-N-ACETYLUMAMINOYL	2.76e+00
26	1013	DNA POLYMERASE (EC 2.7	2.96e+00
27	DPOL_NPVLD	DNA-DIRECTED RNA POLYM	4.90e+00
28	RP06_VACCV	TYPE II RESTRICTION EN	4.90e+00
29	T2N5_NOSST	UDP-N-ACETYLUMAMINOYL	3.81e+00
30	6.6	REPAIR MOUSE	4.90e+00
31	508	INTERFERON-INDUCED GUA	3.81e+00
32	589	BETA-GALACTOSIDASE PRE	3.81e+00
33	669	DNA POLYMERASE (EC 2.7	6.28e+00
34	1235	DNA-DIRECTED RNA POLYM	6.28e+00
35	RP06_VAVV	N-ACETYL-GAMMA-GLUTAM	6.28e+00
36	185	ENOLASE (EC 4.2.1.11)	6.28e+00
37	395	PHOSPHOGlycerate KINAS	6.28e+00
38	416	PHOSPHOGlycerate KINAS	6.28e+00
39	PKC_TRIPE	INTER-ALPHA-TRIPSPIN IN	8.05e+00
40	914	HYPOHETICAL PROTEIN M	1.03e+01
41	1295_METPA	HYPOHETICAL 43.4 KD P	1.03e+01
42	YEIM_ECOLI	HYPOHETICAL 43.4 KD P	1.03e+01
43	MURD_STRAU	UDP-N-ACETYLUMAMINOYL	8.05e+00
44	VPT4_NPVCF	P74 PROTEIN	8.05e+00
45	M8H2_NECPC	DNA MISMATCH REPAIR PR	8.05e+00
46	937	DNA-DIRECTED RNA POLYM	8.05e+00
47	2452	DNA-DIRECTED RNA POLYM	8.05e+00

## ALIGNMENTS

ID	MURC	STAAU	1	STANDARD;	PRT;	437	AA.
AC	031211;						
DT	15-JUL-1998	(REL. 36, CREATED)					
DT	15-JUL-1998	(REL. 36, LAST SEQUENCE UPDATE)					
DT	15-JUL-1998	(REL. 36, LAST ANNOTATION UPDATE)					
DE	UDP-N-ACETYLMURAMATE-L-ALANINE	LIGASE (EC 6.3.2.8) (UDP-N-ACETYLMURAMOYL-L-ALANINE SYNTHETASE).					
GN	MURC.						
OC	STAPHYLOCOCCUS AUREUS.						
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;						
OC	STAPHYLOCOCCUS.						
RN	(1)						
RP	SEQUENCE FROM 'A.A.						
RA	LOWE A.M., DERESIEWICZ R.T.;						
RL	SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.						
CC	-1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).						
CC	-1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL-L-ALANINE -						
CC	ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.						
CC	-1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.						
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).						
CC	-1- SIMILARITY: BELONGS TO THE MORCDEE FAMILY.						
CC	-----						
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration						
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -						
CC	the European Bioinformatics Institute. There are no restrictions on its						
CC	use by non-profit institutions as long as its content is in no way						
CC	modified and this statement is not removed. Usage by and for commercial						
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>						
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).						
CC	-----						
DR	EMBL; AF034076; G2642659; -						
KM	PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;						
KW	ATP-BINDING.						
FT	NP_BIND 108 114 ATP (POTENTIAL).						
SQ	SEQUENCE 437 AA; 49176 MW; 64D68582 CRC32;						
Query Match	97.3%;	Score 1454;	DB 1;	Length 437;			
Best Local Similarity	97.7%;	Pred. No. 1.39e-245;					
Matches 210;	Conservative 3;	Mismatches 2;	Indels 0;	Gaps 0;			
Db 223	FKDSDIYAQNIQITDKGAFDYVYDGEFFDFTLSPOYGDHTVLAALVAIATSYLEKLDV	282					
QY	1 FKDSDIYAQNIQITDKGAVDYVDGEFFDFTLSPOYGDHTVLAALVAIATSYLEKLDV	60					
Db 263	TIKAELETFGGVKRRNETTIANQVYVDVAHHPREISATTEIKAKKYPKEVVAVFOP	342					



QY 61 TNKEALETFEGVKKRNETTIANOVIVDDYAHHPREISATIDARKKYPKREVVAVFOP 120

Db 343 HTFSRTQAFLEFAESLSKADRVFLCEIFGSIRENTGALTIDODLIDKREGASLINEDSIN 402

QY 121 HTFSRTQAFLEFAESLSKADRVFLCEIFGSIRENTGALTIDODLIDKREGASLINEDSIN 180

Db 403 VLEQFDNAVVLFEKGADIQKLOMAYLIDKGMKNAP 437

QY 181 VLEQFDNAVVLFEKGADIQKLOMAYLIDKGMKNAP 215

RESULT 2  
ID MURC\_BACSU STANDARD: PRT: 432 AA.

AC P40778;

DT 01-FEB-1995 (REL. 31, CREATED)

DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE UDP-N-ACETYLUMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-ACETYLUMURAMATE--L-ALANINE SYNTHETASE).

GN MURC.

OS BACILLUS SUBTILIS.

OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;

OC BACILLUS.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 98048467.

RA LAPIDUS A., GALLERON N., SOROKIN A., EHRLICH S.D.;

RT "sequencing and functional annotation of the Bacillus subtilis genes in the 200 kb rnb-dnaB region.";

RL MICROBIOLOGY 143:3431-3441(1997).

RN [2]

RP SEQUENCE OF 85-432 FROM N.A.

RC STRAIN-168 / MARBURG;

RX VARON D., BRODY M.S., PRICE C.W.;

RT "Bacillus subtilis operon under the dual control of the general stress transcription factor sigma B and the sporulation transcription factor sigma H.";

RL MOL. MICROBIOL. 20:339-350(1996).

CC -1- FUNCTION: CELL WALL FORMATION.

CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLUMURAMATE + L-ALANINE -> ADP + ORTHOPHOSPHATE + UDP-N-ACETYLUMURAMATE-L-ALANINE.

CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).

CC -1- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: AF008220; G2293216; -

DR EMBL: L31845; G556014; -

DR SUBTILIS; BG10973; MURC.

KM PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;

KW ATP-BINDING.

FT NP\_BIND 108 114 ATP (POTENTIAL).

SQ SEQUENCE 432 AA: 48364 MW: 163166CB CRC32:

Query Match 57.7%: Score 863; DB 1; Length 432; Best local Similarity 59.1%: Pred. No. 1.20e-132; Matches 120; Conservative 32; Mismatches 51; Indels 0; Gaps 0;

Db 225 EENDFOARNIVKSTEGTFVFVNTFFYDFFYPAYGHNVLSIAVIALCHEEIDSSI 284

QY 3 DSDDIYAOIFQITDGTAVDVYVGEFFYDHTLSPOYGDHFLNLAVALIAYISLEKLVTN 62

Db 285 IKHAKLFGGVKKRNFENKQADYVLDIDYAHHPREIKVTTTEANAKQKPDDEIYAVFOPHT 344

QY 63 IKHAKLFGGVKKRNFENKQADYVLDIDYAHHPREIKVTTTEANAKQKPDDEIYAVFOPHT 122

Db 345 FTPTQOFLEFAESLSGADCVYLDIDIGSARENAGKLTIGLOGKINAKLIEEDOTS 404

QY 123 FSRQAFLEFAESLSGADCVYLDIDIGSARENAGKLTIGLOGKINAKLIEEDOTS 182

Db 405 KAHDAVLIFFMGAGDIQKYNMAY 427

QY 183 EQFDNAVVLFEKGADIQKLOMAY 205

RESULT 3  
ID MURC\_SYNT3 STANDARD: PRT: 505 AA.

AC P74528;

DT 15-JUL-1998 (REL. 36, CREATED)

DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE UDP-N-ACETYLUMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-ACETYLUMURAMATE--L-ALANINE SYNTHETASE).

GN MURC OR SLR1423.

OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).

OC BACTERIA: CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 97061201.

RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y., MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T., HOSONUCHI T., MATSUO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S., SHIMO S., TAKUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M., TABATA S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.";

RL DNA RES. 3:109-136(1996).

CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLUMURAMATE + L-ALANINE -> ADP + ORTHOPHOSPHATE + UDP-N-ACETYLUMURAMATE-L-ALANINE.

CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).

CC -1- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: D90916; G1653723; -

KM PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;

KW ATP-BINDING.

FT NP\_BIND 164 170 ATP (POTENTIAL).

SQ SEQUENCE 505 AA: 54513 MW: B3D7FC7C CRC32:

Query Match 22.9%: Score 343; DB 1; Length 505; Best local Similarity 43.8%: Pred. No. 1.62e-37; Matches 57; Conservative 29; Mismatches 40; Indels 4; Gaps 4;

Db 322 GDHNSNALAAVAVGRLLGIDFPIVIAQIASENGAKRRECKGNGCTFTIDYAHHPSE 381

QY 39 GDHVTNALAVIAISYLEKIDVTNIKEALETFEGVKKRNETTIANOV-IVDDYAHHPRE 97

Db 382 LNTLAAAKQVYHNGYERVAIFOPHRKSRITTEAFATKADKADYVLDIYAGSD 441

QY 98 ISATIDTARKKYPH-R-E-VVAVFOPHTSRTQAFLEFAESLSKADRVFLCEIFGSI 154

Db 442 NPYNIGEDL 451

QY 155 NTGALTIDOL 164

RESULT 4  
ID MURC\_PORGI STANDARD: PRT: 433 AA.



```

AC Q01831:1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE UDP-N-ACETYLMURAMATE--ALANINE LIASE (EC 6.3.2.8) (UDP-N-
DE ACETYLMURAMOYL-L-ALANINE SYNTHETASE).
GN MURC.
OS PORPHYROMONAS GINGIVALIS (BACTERIOIDES GINGIVALIS).
OC BACTERIA: CYTOPHAGALES; BACTEROIDACEAE; PORPHYROMONAS.
RN [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-381;
RX MEDLINE; 96118684.
RA ANSAI T., YAMASHITA Y., AMANO S., SHIBATA Y., WACHI M., NAGAI K.,
RA TAKEHARA T.,
RA "A mucic gene in Porphyromonas gingivalis 381."
RL MICROBIOLOGY 141:2047-2052(1995).
RN [2]
RN RP REVISIONS.
RC STRAIN-381;
RA ANSAI T.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL + L-ALANINE =
CC ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.
CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE MORCDEF FAMILY.
CC -----
CC CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D84504; D1025270; -
KW PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;
KW NP-BIND.
FT NP-BIND 114 120 ATP (POTENTIAL).
SQ SEQUENCE 433 AA; 48309 MW; 38C4714B.CRC32;
Query Match 21.9%; Score 327; DB 1; Length 433;
Best Local Similarity 40.7%; Pred. No. 9,15e-35;
Matches 46; Conservative 32; Mismatches 33; Indels 2; Gaps 2;
Db 261 VENAVAAAMAIHLNVTVEELRSGLASRSGSRREKXULDERVYLIDDYAHHPVEDAA 340
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | |
Oy 43 VLNALAVAIISTLELDVDTNKEALEETGGVKKRRNETTTANQVY-VDYVAHHPREISAT 101
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 341 IRSVEIISGKHINGIFOPHLYSRTADEYDFAFKSLMDQVLLDIPA-RE 392
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | |
Oy 102 IDLAKKKPHKEVYAVFQPHFTRQALNPEABESLSTADRYFLCEITGSGIRE 154
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | |

RESULT 5
ID MURC_SYN7 STANDARD: PRT: 280 AA.
AC P95836;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE UDP-N-ACETYLMURAMATE--ALANINE LIASE (EC 6.3.2.8) (UDP-N-
DE ACETYLMURAMOYL-L-ALANINE SYNTHETASE) (FRAGMENT).
GN MURC.
OS SYNECHOCOCUS SP. (STRAIN PCC 7942) (ANACYSTIS NIDULANS R2).
OC BACTERIA: CYANOBACTERIA; CHROCOCCALES; SYNECHOCOCUS.
RN [1]
RN RP SEQUENCE FROM N.A.
RC KIRZNER S., KAPLAN A.;
RA SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL + L-ALANINE =

```

CC ADP ORNTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.  
CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: CYTOSOLSMIC (PROBABLE).  
CC -1- SIMILARITY: BELONGS TO THE MORCDEF FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: D86147; G1841788; -  
KW PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;  
KW ATP-BINDING..  
FT NON\_TER 1  
SQ SEQUENCE 280 AA; 30753 MW; AE03BP51 CRC32;  
  
Query Match 20.3%; Score 303; DB 1; Length 280;  
Best Local Similarity 40.7%; Pred. No. 1,11e-30;  
Matches 66; Conservative 32; Mismatches 52; Indels 12; Gaps 6;  
  
Db 64 GQAATWEGELGRIHPLMEAHNLNMAALAAACRLG-MDFASIEGLAGFGARR 122  
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Oy 18 GAVDYVDEGEYDFHSPOYGDHVTVALALVALAIS-YLEKIDVYNIXEALTEFGVARR 76  
  
Db 123 FEFRGSAOIGFVDYAHHPSELAATLAARLQIDSGCSRLPEYFKRIYALFQPHRYERT 182  
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Oy 77 FNETIANOV-IVDDYAHHPREISATIDTAR-K-----KPH-KYVAAYFQPHFSRT 126  
  
Db 183 QAFLAFAOSFGPADLVLSIDYAGERNPQOLSGOTLADAI 224  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Oy 127 QAFLNFAESLSKADRVFLCEIFGSRNTGALTITQIDIDKI 168  
  
RESULT 6  
AC MURC\_HAEIN STANDARD; PRT; 475 AA.  
P45066;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE UDP-N-ACETYLMURAMATE-ALANINE LIGASE (EC 6.3.2.8) (UDP-N-  
DE ACETYLMURAMOYL-L-ALANINE SYNTHETASE).  
DE MURC OR H1139.  
GN HAEMOPHILUS INFLUENZAE.  
OS BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURACEAE;  
OC HAEMOPHILUS.  
OC [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-RD / KM20;  
RC MEDLINE: 95350630.  
RX  
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,  
RA KEELAVAGE A.R., BULF C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,  
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOODYNE J.D.,  
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLOBER A., KELLEY J.M.,  
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., REDBLOM E., COTTON M.D.,  
RA WEIDENACK T.R., HANNA M.C., NGUYEN D.T., SAUNDER D.M., BRANDON R.C.,  
RA FINE L.D., FITCHMAN J.L., FUHRMANN J.L., GEORHAGEN N.S.M.,  
RA GINEM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,  
RA VENTER J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT influenzae Rd.";  
RL SCIENCE 269:496-512(1995).  
CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL-L-ALANINE =  
CC ADP + ORNTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.  
CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: CYTOSOLSMIC (PROBABLE).  
CC -1- SIMILARITY: BELONGS TO THE MORCDEF FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL: U32794; G1574695; -  
CC TIGR: H11139; -  
CC PEPTIDOGLYCAN SYNTHESIS: CELL WALL, CELL DIVISION; LIGASE;  
CC ATP-BINDING.  
CC NP\_BIND 125 131 ATP (POTENTIAL).  
CC SEQUENCE 475 AA; 51994 MW; DE40C270 CRC32;

Query Match 19.5%; Score 291; DB 1; Length 475;  
Best Local Similarity 34.6%; Pred. No. 1,176-28;

Matches 62; Conservative 48; Mismatches 51; Indels 18; Gaps 15;

DB 289 GKNMNLATLALVAKKEGIANEALADFOGAGRRDQCEITRPNKGYLVDDYGH 348  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 39 GHTVNLALAVIAISYLEKIDYTNKEALETEGVRKRFNE--TT-I-AN-QV-IVDDYAH 93  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
DB 349 HPEVGVTKAREGMDRIVMIFQPHRSRTDLDFDVQVLSQVADLMVDYVAAE 408  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 94 HREIATITDTARKKYPHKEVAVFQPHFSRTQAFLEAEBSLRADRYFLCEIF--G- 150  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
DB 409 APIVGADSKSLCSIRNL-GKVD-PIIVSDTSQGLVDQIIDQDGLLAOGAGSVSKI 465  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 151 S-IR-ENTGAL--TIQDLIDKIGASGLINEDS-I-NVLEGF--DNMVFVFMGAGDIQKL 201  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
OY

RESULT 7  
ID MURC\_BORBU STANDARD; PRT; 468 AA.  
AC 05157;

DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-

GN MURC OR BB0817.

OS BORRELLIA BURGDORFERI (LYME DISEASE SPIROCHETE).

OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; BORRELLIA.

RA [1]

RA SEQUENCE FROM N.A.

RA STRAIN=ATCC 35210 / B31;

RA MEDLINE: 98065943;

RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,  
RA LATHEGIRA R., WHITE O., KETCHUM K.A., DODSON R., HICKER E.K., GWINN M.,  
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,  
RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,  
RA VAN VUOT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,  
RA UTTERBACK T., WATHEY L., McDONALD L., ARTICH P., BOWMAN C.,  
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,  
RA SMITH H.O., VENTER J.C.;

RA "Genomic sequence of a Lyme disease spirochete, Borrelia  
RA burgdorferi";

RA NATURE 390:580-586(1997).

CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL + L-ALANINE -

CC -1- ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.

CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).

CC -1- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE001180; G2688761; -  
CC TIGR: BB0817; -

KW PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;  
KW ATP-BINDING.  
FT NP\_BIND 121 127 ATP (POTENTIAL).  
SQ SEQUENCE 468 AA; 53480 MW; 09511A39 CRC32;

Query Match 18.9%; Score 283; DB 1; Length 468;  
Best Local Similarity 38.3%; Pred. No. 2,566-27;  
Matches 59; Conservative 33; Mismatches 51; Indels 11; Gaps 8;

DB 316 IKRIAKNGGIRRVYKKEGVYIMDDYAHHPREIKNTLFGINFYKKNRIILDFMPH 375  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 63 IEALETEGVRKRFNE--TTI-AN-QV-IVDDYAHHPREISATITDTARKKYPHKEVAVFQPH 121  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
DB 376 TPTREKPEADVEVLSADILILNINYLSPNENPNDELSTKYLKIRKINRYFFED 435  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 122 TTSRQAFLEAEBSLRADRYFLCEIFGISTEN--TGALTQD-L-IDKIGAS-LINE 176  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
DB 436 VKDSINFKSLISGDLFTMGAGNFTLHD-FL 468  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 177 --DSINVLQFQNAVLF--MGAGDIQKLQNAVYL 206  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
OY

RESULT 8  
ID MURC\_TREPA STANDARD; PRT; 481 AA.  
AC 08361;

DT 15-DEC-1998 (REL. 37, CREATED)

DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-

GN MURC OR TP0341.

OS BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; TREPONEMA.

OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; TREPONEMA.

RA [1]

RA SEQUENCE FROM N.A.

RA STRAIN=NICHOLS;

RA MEDLINE: 98332770.

RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,  
RA DODSON R., GWINN M., HICKER E.K., CLAYTON R., KETCHUM K.A.,  
RA SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,  
RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDEMBERK M., UTTERBACK T.,  
RA McDONALD L., ARTICH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,  
RA HATCH B., HORST K., ROBERTS K., WATHEY L., WEIDMAN J., SMITH H.O.,  
RA VENTER J.C.;

RA "Complete genome sequence of Treponema pallidum, the syphilis  
RA spirochete";

RA SCIENCE 281:375-388(1998).

CC -1- FUNCTION: CELL WALL FORMATION.

CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL + L-ALANINE -

CC -1- ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.

CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).

CC -1- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE001213; G3322616; -

CC TIGR: TP0341; -  
CC PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;  
CC ATP-BINDING.  
FT NP\_BIND 122 128 ATP (POTENTIAL).  
SQ SEQUENCE 481 AA; 53762 MW; 9F643678 CRC32;

Query Match 17.7%; Score 264; DB 1; Length 481;  
Best Local Similarity 34.0%; Pred. No. 3,626-24;  
Matches 50; Conservative 35; Mismatches 55; Indels 7; Gaps 7;



DB 280 GEFYVALPEHSHVNLATGALALSLVKQGEVYEHITLARKVALAQCRSESVLG 339  
| | | | : : : : | | | | : : : : | | | | : : : : | | | | : : : :  
27 GEFYHFLSPQYG-DHT-VLN-ALANIALSYLE-KID-VTNKEALETFGKVRRENE-T 80  
DB 340 EVRGILFMDVYHHPHAIKKTLLGKTFEPERRIVDEMSHVSRTALLTFEAFSFOOA 399  
| | | | : : : : | | | | : : : : | | | | : : : : | | | | : : : :  
81 TANQVYVDVYHHPREISATIDTARKKYPHEVAVAFQPHFSRTQAFLENEAESLSKA 140  
DB 400 DVYILHEIYASAREVYQGEVNGEHLFE 426  
| | | | : : : : | | | | : : : : | | | | : : : : | | | | : : : :  
QY 141 DVFLECEIGSIRE-NTGALTIDDLID 166  
RESULT 9  
ID MURC\_MYCTU STANDARD: PRT: 494 AA.  
AC 006225:  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE UDP-N-ACETYLURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-  
ACETYLURAMOYL-L-ALANINE SYNTHETASE).  
GN MURC OR MTCY270.16.  
OS MYCOBACTERIUM TUBERCULOSIS.  
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;  
OC ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA BADCOCK K., CHURCHER C.M., BARRELL B.G., RAJANDREAN M.A., PARKHILL J.;  
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBI DATA BANKS.  
CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLURAMOYL + L-ALANINE =  
CC ADP + ORTHOPHOSPHATE + UDP-N-ACETYLURAMOYL-L-ALANINE.  
CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
CC -1- SIMILARITY: BELONGS TO THE MURCODE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: 295388; E315954; -  
KW PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;  
KM ATP-BINDING.  
FT NP\_BIND 122 128 ATP (POTENTIAL);  
SQ SEQUENCE 494 AA; 51176 MW; B41D1343 CRC32;  
Query Match 17.3%; Score 259; DB 1; Length 494;  
Best Local Similarity 31.9%; Pred. No. 2,40e-23;  
Matches 60; Conservative 44; Mismatches 69; Indels 15; Gaps 11;  
DB 295 GRHMLNALGALLAAVOICAPADEVDLGLAGEFGVGRFELVGTGCVGRASYRFPDYAH 354  
| | | | : : : : | | | | : : : : | | | | : : : : | | | | : : : :  
QY 39 GGHYVNLALVAIAISYLEKLDVTNKEALETFGKVRKRN--ET-TIANQ-VIV-DDYAH 93  
DB 355 HPTESATIAAARMLVLEQDGGRCVVFQPHLYSRTKAFRAEGRALNADAEVFLDYVG 414  
| | | | : : : : | | | | : : : : | | | | : : : : | | | | : : : :  
QY 94 HPREISATIDTARKKYPKREV--VAVFQPHFSRTQAFLENEAESLSKADRVFLCEIFG 150  
DB 415 A-REOPLAIVSASVAEHYTPKRYVPDSNAVOYAAASPDVIVYMGADVITLIGE 473  
| | | | : : : : | | | | : : : : | | | | : : : : | | | | : : : :  
QY 151 STRENT-GALTIDDLIDKIEGA-SLINE-DSIN-VLEQFD-NAVVLFGAGDIQLOA 204  
DB 474 ITALRVR 481  
| | | | : : : : | | | | : : : : | | | | : : : : | | | | : : : :  
QY 205 YLDKLGK 212  
RESULT 10

ID MURC\_BUCAP STANDARD: PRT: 483 AA.  
AC 051926:  
DT 15-DEC-1998 (REL. 37, CREATED)  
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE UDP-N-ACETYLURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-  
ACETYLURAMOYL-L-ALANINE SYNTHETASE).  
GN MURC.  
OS BUCHNERA APHIDICOLA.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; BUCHNERA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 98087557.  
RA BAUMANN L., BAUMANN P.;  
RT "Characterization of ftsZ, the cell division gene of Buchnera  
RT aphidicola (endosymbiont of aphids) and detection of the product.";  
RL CURR. MICROBIOL. 36:85-89(1998).  
CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLURAMOYL + L-ALANINE =  
CC ADP + ORTHOPHOSPHATE + UDP-N-ACETYLURAMOYL-L-ALANINE.  
CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
CC -1- SIMILARITY: BELONGS TO THE MURCODE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF012886; G2738586; -  
KW PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;  
KM ATP-BINDING.  
FT NP\_BIND 124 130 ATP (POTENTIAL);  
SQ SEQUENCE 483 AA; 54852 MW; 23E8697B CRC32;  
Query Match 16.3%; Score 243; DB 1; Length 483;  
Best Local Similarity 32.8%; Pred. No. 9.72e-21;  
Matches 60; Conservative 49; Mismatches 56; Indels 18; Gaps 13;  
DB 289 GRHMLNLTAIAFAFTYQKIDPEKIKYOSLKNKFGSRFEVYGLFKIKKSIONKSMVL 348  
| | | | : : : : | | | | : : : : | | | | : : : : | | | | : : : :  
QY 39 GGHYVNLALVAIAISYLEKLDVTNKEALETF-GGKVR-----R-F-NETTANQ-VI-V 88  
DB 349 NDYGHHPTELSETIKTKRSWPKKMLIMFQPHRYTRTMYFDFIKLSQVSLILNV 408  
| | | | : : : : | | | | : : : : | | | | : : : : | | | | : : : :  
QY 89 DDYAHHPREISATIDTARKKYPKREVAVFQPHFSRTQAFLENEAESLSKADRVFLCEI 148  
DB 409 YSANSFISGASFSLSYDIKKIKKRVILVTRNLLVTLPHLNGNDIILIQAGDIDT 468  
| | | | : : : : | | | | : : : : | | | | : : : : | | | | : : : :  
QY 149 FESIRE-NTGA-LTIQDLIDIE-GASLI-NEDSI-N-VLEQFD-NAVVLFGAGDIQOK 200  
DB 469 IIN 471  
| | | | : : : : | | | | : : : : | | | | : : : : | | | | : : : :  
QY 201 ION 203  
RESULT 11  
ID MURC\_HELPY STANDARD: PRT: 449 AA.  
AC 025340:  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE UDP-N-ACETYLURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-  
ACETYLURAMOYL-L-ALANINE SYNTHETASE).  
GN MURC OR HP0623.  
OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).  
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; HELICOBACTER GROUP;  
OC HELICOBACTER.  
RN [1]  
RP SEQUENCE FROM N.A.



RC STRAIN=26695 / ATCC 700392;  
 RX MEDLINE; 97394467.  
 RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,  
 RA FLEISCHMANN R.D., KETCHUM K.A., KLENN H.-P., GILL S., DOUGHERTY B.A.,  
 RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,  
 RA LOTTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLIDER A.,  
 RA MCENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,  
 RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,  
 RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,  
 RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,  
 RA VENTER J.C.:  
 RT "The complete genome sequence of the gastric pathogen *Helicobacter*  
 RT *pylori*.";  
 RL NATURE 388:539-547(1997).  
 CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMYL + L-ALANINE =  
 CC ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMYL-L-ALANINE.  
 CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000576; G2313743; -.  
 DR TIGR; HP0623; -.  
 KM PEPTIDOGLYCAN SYNTHESIS: CELL WALL; CELL DIVISION; LIGASE;  
 KW ATP-BINDING.  
 FT NP\_BIND 121 127 ATP (POTENTIAL).  
 SO SEQUENCE 449 AA; 50793 MW; EC3EDD2A CRC32;  
 Query Match 15.7%; Score 235; DB 1; Length 449;  
 Best Local Similarity 29.9%; Pred. No. 1.89e-19;  
 Matches 46; Conservative 40; Mismatches 60; Indels 8; Gaps 7;  
 Db 235 EKDLYNIQYILKD-GEPTYSFELKD-LGALLWGLGSHNATNA-STALISALDELHEE 291  
 QY 3 DSDIYAFQITKGTAVDYVDEGFDFHLSPOYGHYTLNMAVAIYSYLEKLDVTN 62  
 Db 292 IRNNLNFKGGKRRDILQKNAIILIDYPAHPTISATLSARIRYALNTQEKIIVM 351  
 QY 63 IKEALETFGSKRRFNETTIANQYIVDYAHHPREISATIDTARK--KYPHK-E-VVAVF 118  
 Db 352 QAHKYSRLMDNLEEFKCFSEHCDRLILPYSA 385  
 QY 119 QPHFSTRQAFNLNFAESLSKA-DRVFLCEIFGS 151  
 RESULT 12  
 ID MURC\_ECOLI STANDARD: PRT: 491 AA.  
 AC P17952; 007099;  
 DT 01-NOV-1990 (REL. 16, CREATED)  
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-  
 DE ACETYLMURAMYL-L-ALANINE SYNTHETASE).  
 GN MURC.  
 OS ESCHERICHIA COLI.  
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
 OC ESCHERICHIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE; 90326550.  
 RA IKEDA M., MACHI M., JUNG H.K., ISHINO F., MATSUHASHI M.:  
 RT "Nucleotide sequence involving *murG* and *murC* in the *mra* gene cluster  
 RT region of *Escherichia coli*.";  
 RL NUCLEIC ACIDS RES. 18:4014-4014(1990).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE; 92334977.  
 RA YURA T., MORI H., NAGAI H., NAGATA T., ISHIHAMA A., FUJITA N.,  
 RA ISONO K., MIZOBUCHI K., NAKATA A.:  
 RT "Systematic sequencing of the *Escherichia coli* genome: analysis of  
 RT the 0-2.4 min region.";  
 RL NUCLEIC ACIDS RES. 20:3305-3308(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE; 97426617.  
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
 RA RILEY M., COLINDO-YRDES J., GLASNER F.D., ROSE C.K., MAYHEW G.F.,  
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,  
 RA MAU B., SHAO Y.:  
 RT "The complete genome sequence of *Escherichia coli* K-12.";  
 RL SCIENCE 277:1453-1474(1997).  
 RN [4]  
 RP MUTANT MURC3, SEQUENCE FROM N.A.  
 RC STRAIN-CGSC 5988;  
 RX MEDLINE; 97309380.  
 RA EVELAND S.S., POMPLIANO D.L., ANDERSON M.S.:  
 RT "Conditionally lethal *Escherichia coli* murein mutants contain point  
 RT defects that map to regions conserved among murein and folsyl  
 RT poly-gamma-glutamate ligases: identification of a ligase  
 RT superfamily.";  
 RL BIOCHEMISTRY 36:6223-6229(1997).  
 RN [5]  
 RP SEQUENCE OF I-14, AND CHARACTERIZATION.  
 RX MEDLINE; 95324553.  
 RA LIGER D., MASSON A., BLANOT D., VAN HERGENOORT J., PARQUET C.:  
 RT "Over-production, purification and properties of the  
 RT uridine-diphosphate-N-acetylmuramate-L-alanine ligase from  
 RT *Escherichia coli*.";  
 RL EUR. J. BIOCHEM. 230:80-87(1995).  
 CC -1- FUNCTION: CELL WALL FORMATION.  
 CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMYL + L-ALANINE =  
 CC ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMYL-L-ALANINE.  
 CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X52644; G42056; -.  
 DR EMBL; X55034; G40859; -.  
 DR EMBL; D10483; G216505; -.  
 DR EMBL; AE000118; G1786279; -.  
 DR EMBL; U67892; G2177094; -.  
 DR PTR; J00545; CECCAM.  
 DR PTR; S40601; S40601.  
 DR ECGENE; EG10619; MURC.  
 KW PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;  
 KM ATP-BINDING.  
 FT NP\_BIND 126 132 ATP (POTENTIAL).  
 FT MURAGEN 344 G->D: IN MURC3.  
 SO SEQUENCE 491 AA; 53626 MW; C68BD87D CRC32;  
 Query Match 15.5%; Score 232; DB 1; Length 491;  
 Best Local Similarity 38.8%; Pred. No. 5.72e-19;  
 Matches 47; Conservative 28; Mismatches 36; Indels 10; Gaps 5;  
 Db 290 GRNNAALAAVAVATEEGIDDEALIRALESFQGTGRREDFLGEPLEPYNGSKGTAMV 349  
 QY 39 GDBTVNALAVIAISYLEKIDVTINIKALETFGCVKRRF---E-T-TI--AN-QVIV 88



Db	350	DGCGHPEVATVITKAAGWPMKLVLMFOPHRTFRDLTYDDPCANVLQVDLLMEY	409
Oy	89	DDIAHPRELISATIDANKKTPHKEVAVFOPHTSRIOAFLEPESLSRADRVLCFI	148
Dd	410	y 410	
Oy	149	f 149	
RESULT	13		
ID	MPL_ECOLI	STANDARD:	PRT; 457 AA.
AC	P376804;		
DT	01-OCT-1994 (REL. 30, CREATED)		
DT	15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)		
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)		
DE	UDP-N-ACETYLMURAMATE:L-ALANYL-GAMMA-D-GLUTAMYL-MESO-DIAMINOPIMELATE		
GN	MPL.		
OS	ESCHERICHIA COLI.		
OC	BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERiaceae;		
CC	ESCHERICHA.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-K12 / MG1655;		
RX	MEDLINE; 95334362.		
RA	BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,		
RT	BLATTNER F.R.;		
RL	"Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.8 through 100 minutes.";		
RN	Nucleic Acids Res. 23:2105-2119.(1995).		
RP	[2]		
RX	SEQUENCE OF 1-64 FROM N.A.		
RA	MEDLINE; 88335617.		
RT	HAMILTON W.D.O., HARRISON D.A., DYER T.A.;		
RL	"Sequence of the Escherichia coli fructose-1,6-bisphosphatase gene.";		
RN	Nucleic Acids Res. 16:8707-8707.(1988).		
RN	[3]		
RP	IDENTIFICATION.		
RX	MEDLINE; 95075659.		
RT	BORODOVSKY M., RUDD K.E., KOONIN E.V.;		
RL	"Intrinsic and extrinsic approaches for detecting genes in a bacterial genome.";		
RN	Nucleic Acids Res. 22:4756-4767.(1994).		
RP	[4]		
RX	CHARACTERIZATION.		
RT	MEDLINE; 96404780.		
RA	MENGIN-LECREUX D., VAN HEIJENOORT J., PARK J.T.;		
RT	"Identification of the mpl gene encoding UDP-N-acetylmuramate: l-alanyl-gamma-D-glutamyl-meso-diaminopimelate lyase in Escherichia coli and its role in recycling of cell wall peptidoglycan.";		
RL	J. Bacteriol. 178:5347-5352.(1996).		
CC	- I - FUNCTION: INVOLVED IN CELL WALL FORMATION. REUTILIZES THE INTACT TRIPEPTIDE L-ALANYL-GAMMA-D-GLUTAMYL-MESO-DIAMINOPIMELATE. - I - SIMILARITY: BELONGS TO THE MORCEP FAMILY.		
CC	- I -		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb.sib.ch).		
DR	EMBL; U14003; G537075; -;		
DR	EMBL; AE000494; G1790680; -;		
DR	EMBL; X12545; -; NOT_ANNOTATED_CDS.		
DR	ECOGENE; EG12440; MPL.		
DR	HSSP; P14900; IUDG.		
KW	PEPTIDOGLYCIN SYNTHESIS; CELL DIVISION; CELL WALL; LIGASE;		
KW	ATP-BINDING.		
FT	NP_BIND 110 116 ATP (POTENTIAL).		
SO	SEQUENCE 457 AA; 49874 MW; AABF8001 CRC32;		

[illegible]



DB 372 KDIAPALGRADAVEM 387  
OY 130 LNEFAESLSKADRVFL 145

Search completed: Wed Nov 24 00:45:09 1999  
Job time : 33 secs.

RESULT 15  
ID MORE\_BACSU STANDARD: PRT: 494 AA.  
AC Q03523;  
DT 01-OCT-1993 (REL. 27, CREATED)  
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE UDP-N-ACETYLMURAMOIDALNYL-D-GLUTAMATE--2,6-DIAMINOPIMELATE LIGASE  
DE (EC 6.3.2.13) (UDP-N-ACETYLMURAMYL-TRIPETIDE SYNTHETASE).  
GN MORE.  
OS BACILLUS SUBTILIS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;  
OC BACILLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-168;  
RX MEDLINE: 93171879.  
RA DANIEL R.A., ERRINGTON J.;  
RT "DNA sequence of the mre-murD region of Bacillus subtilis 168.";  
RL J. GEN. MICROBIOL. 139:361-370(1993).  
RN [2]  
RP SEQUENCE OF 1-16 FROM N.A.  
RC STRAIN-168;  
RX MEDLINE: 94118264.  
RA DANIEL R.A., DRAKE S., BUCHANAN C.E., SCHOLLE R., ERRINGTON J.;  
RT "The Bacillus subtilis spvD gene encodes a mother-cell-specific  
penicillin-binding protein required for spore morphogenesis.";  
RL J. MOL. BIOL. 235:209-220(1994).  
CC -1- FUNCTION: CELL WALL FORMATION. DIAMINOPIMELIC ACID ADDING ENZYME.  
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMYL-L-ALANYL-D-  
+ UDP-N-ACETYLMURAMYL-L-ALANYL-D-GLUTAMYL-MESO-2,6-  
DIAMINOPIMELATE.  
CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
CC -1- SIMILARITY: BELONGS TO THE MRCDEE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL: Z15056; G40162; -  
CC EMBL: Z25865; G397896; -  
CC EMBL: Z99111; E1185108; -  
CC PIR: S23914; S23914.  
CC PIR: B47691; B47691.  
CC SUBTILIS; BG10223; MORE.  
CC PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;  
CC ATP-BINDING.  
CC  
CC NP\_BIND 109 115 ATP (POTENTIAL).  
CC FT SEQUENCE 494 AA; 54325 MW; BCC9A893 CRC32;  
CC SQ

Query Match 8.8%; Score 132; DB 1; Length 494;  
Best Local Similarity 28.8%; Pred. No. 3.76e-04;  
Matches 34; Conservative 22; Mismatches 60; Indels 2; Gaps 2;

DB 262 IKNDADVAKNISTACTSDL-VTNKGTKHITMSLVGQFNVYVLAQVATCINAGIPF 320  
OY 1 FKDDSDIYAQIFQJLTDKGTAVDVYDGEFFYDHFSLPQYGDHTVNALAVIAISYLEKLDV 60  
DB 321 EITEAVDELGVGRGFLVNOQGEFFVIVDYAHTPDSENVLETGRDMEGKLEVVV 378  
OY 61 TNKEALTEFGVGRKRFNETITANOV-IVDDYAHHPREISATIDARKKYPKEVVAV 117















Db 418 VLSVLENEIEIKV-DVKSPEDEGIR-VAEVLADGTSFKGAVI-Y-GLGVRKKE 467  
 148 IFGSIKRENTGALTIQDIDKIEGASLINESINLEQFDNAVVLFGMGAGDIQKLT 201

RESULT 9  
 ID P94335 PRELIMINARY; PRT: 90 AA.

AC 01-MAY-1997 (TREMBLREL. 03, CREATED)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE MURC (UDP-N-ACETYLURAMATE--ALANINE LIGASE) (EC 6.3.2.8)  
 DE (UDP-N-ACETYLURAMATE--L-ALANINE SYNTHETASE) (FRAGMENT).  
 GN MURC.

OS CORVNEBACTERIUM GLUTAMICUM.  
 OC BACTERIA: FIRMICUTES: ACTINOBACTERIA: ACTINOBACTERIDAE;  
 OC ACTINOMYCETALES: CORVNEBACTERINAE; CORVNEBACTERIACEAE;  
 OC CORVNEBACTERIUM.

RN 11  
 RP SEQUENCE FROM N.A.

RX MEDLINE: 97382442.  
 RA KOBAYASHI M., ASAI Y., HATAKEYAMA K., KIJIMA N., WACHI M., NAGAI K.,  
 YUKAWA H.,

RT Cloning, sequencing, and characterization of the ftsZ gene from  
 RT Corynebacterium bacteria.  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 236:383-388(1997).

RN 12  
 RP SEQUENCE OF 4-90 FROM N.A.

RA MARCOS P.H., GIL J.A.;  
 RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLURAMATE + L-ALANINE = ADP  
 CC + PHOSPHATE + UDP-N-ACETYLURAMATE + L-ALANINE.

DR EMBL: AB003132; D1022538;  
 DR EMBL: Y08964; E283098;

KM LIGASE.

FT NON\_TER

SQ SEQUENCE 90 AA; 9606 MW; FICD437F CRC32;

Query Match 7.0%; Score 105; DB 2; Length 90;  
 Best Local Similarity 35.2%; Pred. No. 7.0le-01;  
 Matches 31; Conservative 18; Mismatches 30; Indels 9; Gaps 6;

Db 2 FQEFEEALSLADAAVYL-ETIGA-REQPYDVSSSEITDAMTIPVVEPNESAVPERIA 59

OY 129 FLNEFEESISKAPR-VLELIEFGSIRENTGALTIOIDLIDKIEGASLINESIN-VLEQF- 185

Db 60 ELAGPNDIVLTGAGSVTMLAPETLDOL 87

OY 186 D-----NAVLEMGAGDIQKLOKNAVYLDKL 209

RESULT 10  
 ID 069290 PRELIMINARY; PRT: 350 AA.

AC 069290;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)

DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)

DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE UDP-N-ACETYLURAMATE--D-GLUTAMATE--2,6-DIAMINOPIMELATE LIGASE  
 DE (FRAGMENT).

GN MURE.

OS CAMPYLOBACTER JEJUNI.

OC BACTERIA: PROTEOBACTERIA: EPSILON SUBDIVISION; CAMPYLOBACTER GROUP;  
 OC CAMPYLOBACTER.

RN 11  
 RP SEQUENCE FROM N.A.

RC STRAIN-NCIC 11168;  
 RA GRIFFITHS P.L., CONNERTON I.F.;

RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: Y16882; E126307;  
 KM LIGASE.

FT NON\_TER

SQ SEQUENCE 350 AA; 38656 MW; 0EAD2922 CRC32;

Query Match 7.0%; Score 104; DB 2; Length 350;  
 Best Local Similarity 28.7%; Pred. No. 9.22e-01;  
 Matches 31; Conservative 25; Mismatches 45; Indels 7; Gaps 6;

Db 232 ENPALQIQAISL-EGISITVTKNOTF-HIDSPLGLFNLVLVASACNELVPRDI 289

OY 3 DSDDIYA-QIFQITDKTADVVDGEFYDHELSPOYGDHTVINALAVIA-LSYLEKLV 60

Db 290 KDEKASISGCGVCGRVEO--VAKGIYV-FAHTPGIEKVLDTLNK 334

OY 61 TNKEKLETFGGVKKRPNETTIANOVYVDYAHHPREISATIDTARKK 108

RESULT 11  
 ID 062800 PRELIMINARY; PRT: 662 AA.

AC 062800;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)

DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)

DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE) (FRAGMENT).

GN GLBI.

OS CANIS FAMILIARIS (DOG).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC CARNIVORA; FISSIPEDIA; CANIDAE; CANIS.

RN 11  
 RP SEQUENCE FROM N.A.

RC STRAIN-BEAGLE: TISSUE-BRAIN.

RA SMITH B.F., GEORGEON M., BAKER H.J.;

RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING  
 CC BETA-D-GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.

DR EMBL: AF056084; G3025876;  
 DR PROSITE: PS01182; GLYCOSYL\_HYDROL\_F35; 1.

KM HYDROLASE.

FT NON\_TER

FT ACT\_SITE

SQ SEQUENCE 183 183

PROTON DONOR (BY SIMILARITY).

SEQUENCE 662 AA; 74394 MW; 929FDD75 CRC32;

Query Match 7.0%; Score 105; DB 6; Length 662;  
 Best Local Similarity 24.0%; Pred. No. 7.0le-01;  
 Matches 30; Conservative 29; Mismatches 35; Indels 8; Gaps 8;

Db 237 FQFGANI-TAAFOIQKSPKPLVSEFYTGMD-HMGQPSHYRT-EVVASLHDIA 293

OY 1 FQSDDIYAQIQTIDKXGAVDYVDGEFYDHELSPOYGD-HTVINALAVIAISYLEKID 59

Db 294 -HGANYLYMFQGTNFAVYNGANMPYQAPTSYDYDAP-LSAGDLTEKYFALREVIRK 351

OY 60 VTNIKEALETF-GGVKRRF-NETTIANOVYVDYAHHPREISATIDTARKKYPHREYAV 117

Db 352 FE 353

OY 118 FQ 119

RESULT 12  
 ID 069276 PRELIMINARY; PRT: 1161 AA.

AC 069276;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)

DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)

DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE NONTXIC-NONHEMAGGLUTININ.

OS CLOSTRIDIUM BOTULINUM.  
 OC BACTERIA: FIRMICUTES: BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;  
 OC CLOSTRIDIUM.  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NCIC 2916;  
 RA RODRIGUEZ JOVITA M., COLLINS M.D., EAST A.K.;



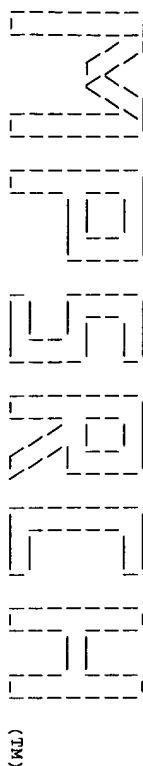




Oy	60	VTNIKEALE-TFGVYKRRNETTIANQVIYDDYAHHPREISATIDTARKKYPHKEVAVF	118
Db	758	QNIYPKFISEMEOCINNINKNTREI-QKJNITENEKLOLINONIFSSLDPFL-NIE	815
Oy	119	OPHTSRTOAFINERFRESLSKADRVFLCEIFGSIRENTG-ALTIDLDIKIEGASLINED	177
Db	816	NKSL-FNSETGLI	829
Oy	178	SINVLQFDNAVLFM	193

Search completed: Wed Nov 24 00:47:43 1999  
 Job time : 138 secs.



\*\*\*\*\*  
  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_tpn n.a. n.a. Smith-Waterman search, using a protein query  
 which has been backtranslated into n.a. using IUPAC symbols

Run on: Sat Nov 27 09:58:57 1999; MasPar time 1136.43 seconds  
 1329.899 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-4  
 Description: (1-215) from US09103287.pep  
 Perfect Score: 2663  
 N.A. Sequence: 1 TTYAARGAAGWNSGAYGAYAT.....TNGGNATGAAARAAGCCTTY 645  
 Comp: AATTCTTCTGNSCTGCTCTTA.....ANCCNACTTCTTTCGNAAR

Scoring table: TABLE bktranslates2  
 Gap: 30

Match STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: emb1-est58  
 1:em\_est10 2:em\_est11 3:em\_est17 4:em\_est18 5:em\_est2  
 6:em\_est19 7:em\_gss1

Database: genbank-est11  
 8:gb\_est1 9:gb\_est10 10:gb\_est11 11:gb\_est12 12:gb\_est13  
 13:gb\_est14 14:gb\_est15 15:gb\_est16 16:gb\_est17  
 17:gb\_est18 18:gb\_est19 19:gb\_est20 20:gb\_est21  
 21:gb\_est22 22:gb\_est23 23:gb\_est24 24:gb\_est25  
 25:gb\_est26 26:gb\_est27 27:gb\_est28 28:gb\_est29  
 29:gb\_est30 30:gb\_est31 31:gb\_est32 32:gb\_est33  
 33:gb\_est34 34:gb\_est35 35:gb\_est36 36:gb\_est37  
 37:gb\_est38 38:gb\_est39 39:gb\_gss3 40:gb\_gss4  
 41:gb\_gss5 42:gb\_gss6

Statistics: Mean 66.667; Variance 79.339; scale 0.840

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	167	6.3	1144	37	B08906 F21H10-T7 IGF Arabidop	1.07e-09
2	162	6.1	2275	20	AF034173 AF034173 Human mRNA (T	1.36e-08
3	158	5.9	398	14	AA486130 ab14b03.r1 Stratagene	1.01e-07
4	156	5.9	576	37	FR0029201 Fugu rubripes GSS sequ	2.71e-07
5	153	5.7	468	41	AQ304744 HS_2190_A2-CO3-T7 CIT	1.17e-06
6	152	5.7	605	36	AA110915 mm2c04.r1 Stratagene	1.91e-06
7	148	5.6	419	34	W29681 mc07e04.r1 Soares mous	1.29e-05
8	148	5.6	455	13	AA423163 ve36b02.r1 Soares mous	1.29e-05
9	148	5.6	459	33	W13700 ma82f05.r1 Soares mous	1.29e-05

C	10	148	5.6	505	36	AA097556	mo02c08.r1 Stratagene	1.29e-05
C	11	148	5.6	515	18	AA839460	vw50c12.r1 Soares mous	1.29e-05
C	12	148	5.6	522	21	A1019831	ua91a06.r1 Soares mous	1.29e-05
C	13	148	5.6	539	10	AA240493	mw33a01.r1 GuayWoodfor	1.29e-05
C	14	148	5.6	542	14	AA547287	vk27h11.r1 Soares mous	1.29e-05
C	15	148	5.6	547	34	W35762	mc14b03.r1 Soares mous	1.29e-05
C	16	148	5.6	553	20	AA895646	va47h01.r1 Stratagene	1.29e-05
C	17	148	5.6	567	13	AA433016	vd94d11.r1 Soares mous	1.29e-05
C	18	148	5.6	568	14	AA520785	T9ESTz26a09.r1 TGME49	1.29e-05
C	19	148	5.6	621	20	AA869547	wq09a05.r1 Bartstead st	1.29e-05
C	20	148	5.6	649	35	AA014494	mc67e03.r1 Soares mous	1.29e-05
C	21	148	5.6	701	23	A1122277	uc60f12.r1 Soares mous	1.29e-05
C	22	148	5.6	747	25	A1317587	uk61a02.r1 Schiller mo	1.29e-05
C	23	148	5.5	999	34	W44178	mc69a05.r1 Soares mous	1.29e-05
C	24	147	5.5	369	34	W52583	z691a04.r1 Pancreatic	2.07e-05
C	25	146	5.5	439	19	R01348	ye81c01.r1 Soares feta	3.31e-05
C	26	146	5.5	474	15	AA573822	nk08a07.r1 NC1_CGAP-Co	3.31e-05
C	27	146	5.5	503	24	A1207935	ap14e03.r1 Schiller ol	3.31e-05
C	28	143	5.4	418	38	AO054129	CIT-HSP-234D9.TF CIT	1.34e-04
C	29	143	5.4	422	37	B47294	HS-1056-A2-E01-MF.abi	1.34e-04
C	30	143	5.4	431	30	R49989	vt56h02.r1 Soares brea	1.34e-04
C	31	145	5.4	451	42	AO439405	HS_5086.r2.C10.SP6 RP	5.29e-05
C	32	144	5.4	576	28	A1542321	S008c09.Sprtime SD Dros	8.42e-05
C	33	145	5.4	577	41	AQ285756	RPC111-9817.TV RPC111	5.29e-05
C	34	143	5.4	596	35	W91238	mt72e12.r1 Soares mous	1.34e-04
C	35	145	5.4	667	39	AO014005	RPC111-25C19.TKBR RPC1	5.29e-05
C	36	145	5.4	669	41	AO373870	RPC111-145B17.TV RPC11	5.29e-05
C	37	145	5.4	2275	20	AF034173	AF034173 Human mRNA (T	5.29e-05
C	38	140	5.3	300	15	C30721	C30721 Yuji Kohara unsp	5.27e-04
C	39	140	5.3	325	12	AA380642	EST93627 Supt cells Ho	5.27e-04
C	40	140	5.3	333	32	D65156	CELK132A7R Yuji Kohara	5.27e-04
C	41	140	5.3	388	38	AO064726	HS_2214.B1.C08.MF CIT	5.27e-04
C	42	141	5.3	418	40	AQ167411	HS_3151.B1.B06.MF CIT	3.35e-04
C	43	141	5.3	535	41	AQ304156	HS_3208.r1.F09.T7 CIT	3.35e-04
C	44	140	5.3	630	41	AQ367690	tox0002006r CGGI Toma	5.27e-04
C	45	139	5.2	345	16	AA646397	vt14h07.r1 Stratagene	8.28e-04

## ALIGNMENTS

RESULT	LOCUS	LENGTH	DNA	GSS	DATE
1	B08906	1144 bp	DNA	GSS	13-MAY-1997
LOCUS	F21H10-T7 IGF Arabidopsis thaliana genomic clone F21H10, genomic				
DEFINITION	B08906				
ACCESSION	B08906				
VERSION	g2090036				
KEYWORDS	B08906.1 GI:2090036				
SOURCE	GSS.				
ORGANISM	thale cress.				
REFERENCE	Arabidopsis thaliana				
AUTHORS	Eukaryotae: mitochondrial eukaryotes; Viridiplantae;				
TITLE	Charophyta/Embryophyta group; Embryophyta; vascular plants; seed				
JOURNAL	plants; Magnoliophyta; Magnoliopsida; Capparales; Brassicaceae;				
COMMENT	Arabidopsis.				
	1 (bases 1 to 1144)				
	Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and				
	Ecker,J.				
	BAC End Sequences at ATCC				
	Unpublished (1997)				
	Other_GSSs: F21H10-SP6				
	Contact: Ecker J.				
	Arabidopsis Thaliana Genome Center				
	University of Pennsylvania				
	Dept. of Biology, University of Pennsylvania, Philadelphia, PA				
	19104				
	Tel: 215-898-9384				
	Fax: 215-898-8780				
	Email: jecker@atgenome.bio.upenn.edu				
	Seg primer: T7				
	Class: BAC ends				
	High quality sequence start: 113				
	High quality sequence stop: 777.				
	Location/Qualifiers				

## FEATURES







ACCESSION	AL025572	sequence.
NID	63262915	
VERSION	AL025572.1	GI:3262915
KEYWORDS	GSS: genome survey sequence.	
SOURCE	Fugu rubripes.	
ORGANISM	Fugu rubripes	
REFERENCE	Eukaryote: Metazoa; Chordata: Vertebrata; Actinopterygii: Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorphia	
AUTHORS	1 (bases 1 to 576) Elgar,G., Clark,M., Smith,S., Week,S., Warner,S., Umrania,Y., Williams,G. and Brenner,S.	
TITLE	Direct Submission	
JOURNAL	Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB, UK. Email: biohelp@hgmrc.mrc.ac.uk	
COMMENT	Vector: pBluescript II KS V-type: phagemid PRIMER: KS DESCR: One pass dye-terminator sequencing of cosmid cloned genomic sequence.	
FEATURES	location/Qualifiers	
source	1..576 /organism="Fugu rubripes" /db_xref="taxon:31033" /clone_1b="cosmid 038C04" /clone="038C04ae2"	
BASE COUNT	169 a 120 c 124 g 157 t 6 others	
ORIGIN		
Query Match	5.98, Score 156, DB 37, Length 576; Best Local Similarity 43.0%; Pred.No. 2,71e-07; Matches 34; Conservative 18; Mismatches 26; Indels 1; Gaps 1;	
Db	234 AAAAGCATCCACGACGAGAGTAGATTGGATTTCACACTAGAT-CCCTGCTACAGTCA 292	
Oy	319 AARARATYCCNCAYARAGARGTNGTCNGCTTTTCARCNCAYACNTTYSNMGNACN 378	
Db	293 CTGGCTTTTCATCTGTAT 311	
Oy	379 CARGCNTTYYTNAAVGART 397	
RESULT	5	
LOCUS	AQ304744 468 bp DNA GSS 15-DEC-1998	
DEFINITION	HS-2190 A2.C03.T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-2190 Col-6 Row-E, genomic survey sequence.	
ACCESSION	AQ304744	
NID	64024530	
VERSION	AQ304744.1	GI:4024530
KEYWORDS	GSS.	
SOURCE	human.	
ORGANISM	Homo sapiens.	
REFERENCE	Eukaryote: Metazoa; Chordata: Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 468) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J., Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.	
TITLE	Construction of a Characterized Clone Resource for Genomic Sequencing	
JOURNAL	Unpublished (1998)	
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector	

FEATURES	source	1. .468	/organism="Homo sapiens"	
			/note="Organ: sperm; Vector: pBelOAc11; BAC Clones In	
			E-Coli DH10B"	
			/db_xref="taxon:9606"	
			/clone="Plate-2190 Col-6 Row-E"	
			/clone_id="CIT Approved Human Genomic Sperm Library D"	
			/sex="male"	
BASE COUNT	157 a	88 c	82 g	140 t 1 others
ORIGIN				
Query Match	5.7%	Score 153;	DB 41;	Length 468;
Best Local Similarity	37.8%;	Pred. No. 1.17e-06;		
Matches	37; Conservative	20; Mismatches	41; Indels	0; Gaps 0;
Db	242	ATTGGACCACTATTATTATTCATTGTCATTTGCGTGTATATATACATAGGTGGGAAC	301	
Qy	113	AYGGAGCAATACNTNTNAAAGCNYGNGNGTATGACNATHWSNTAYTYNGARAAR	172	
Db	302	TACATCTAACATCATTTGAGAGAGCGTTTATATCTT	339	
Qy	173	TNGAVGTACNAAATYTHAARGARGCNTYNGARACNTTY	210	
RESULT	6			
LOCUS	AA10915	605 bp	mRNA	EST 04-FEB-1997
DEFINITION	mm02c04.r1 Strata gene mouse kidney (#937315) Mus musculus CDNA			
	clone IMAGE:20326 5' similar to gb:J04933 Mouse gelsolin gene,			
	complete cds (MOUSE);, mRNA sequence.			
ACCESSION	AA10915			
NID	q1662791			
VERSION	AA10915.1	GI:1662791		
KEYWORDS	EST.			
ORGANISM	house mouse.			
SOURCE	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 605)			
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,			
	Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,			
	Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,			
	Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and			
	Waterston,R.			
	The WashU-HMI Mouse EST Project			
	Unpublished (1996)			
TITLE	On Apr 14, 1993 this sequence version replaced gi:429368.			
JOURNAL				
COMMENT	Contact: Marra M/Mouse EST Project			
	WashU-HMI Mouse EST Project			
	Washington University School of Medicine			
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108			
	Tel: 314 286 1800			
	Fax: 314 286 1810			
	Email: mousestewartson.wustl.edu			
	This clone is available royalty-free through LNL; contact the			
	IMAGE Consortium (infoimage.lnl.gov) for further information.			
	MGI:314174			
	Sed primer: -28ml3 rev1 ET from Amersham			
	High quality sequence stop: 375.			
FEATURES	Location/Qualifiers			
source	1. .605			
	/organism="Mus musculus"			
	/strain="C57/Bl6"			
	/note="Organ: kidney; Vector: pBluescript SK-; Site:1:			
	Ecot1; Site:2: Xho1; Cloned unidirectionally. Primer:			
	Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector;			
	-5' adaptor sequence: 5' GATTTCGGACACAG 3' -3' adaptor			
	sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'.			
	/db_xref="taxon:10090"			



```

/mape="21"
/clone="IMAGE:520326"
/clone_lib="Stratagene mouse kidney (#937315)"
/sex="females"
/tissue_type="kidney"
/dev_stage="4 weeks"
/lab_host="SOLR (kanamycin resistant)"
BASE COUNT      153 a      161 c      144 g      146 t      1 others
ORIGIN

Query Match      5.7%; Score 152; DB 36; Length 605;
Best Local Similarity 39.1%; Pred. No. 1,91e-06;
Matches 34; Conservative 18; Mismatches 35; Indels 0; Gaps 0;

Db 443 ACCGAGACAAACACTCGTCCAGGTGTCAGAGACATACATCATTAACGAGCTT 502
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 326 AYCNCACAAARAGRTNGCNGTNTTTCARCNCACATNTTWSMGNACGACGCT 385
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 386 TTYTNAAGARTTTCGNCAGRMSNTYTW 412
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
LOCUS W29681 419 bp mRNA EST 11-SEP-1996
DEFINITION mc27e04.1 Soares mouse p3MMP19.5 Mus musculus cDNA clone
IMAGE:347838 5' similar to gb:J04953 Mouse gelsolin gene, complete
cds (MOUSE);, mRNA sequence.
ACCESSION W29681
NID g1309830
VERSION W29681.1 GI:1309830
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 419)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMNI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Apr 14, 1993 this sequence version replaced gi:785868.

Contact: Marra M/Mouse EST Project
WashU-HMNI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MG1:219638
Seq primer: EYPrimer
High quality sequence stop: 403.
Location/Qualifiers
1..419
/organism="Mus musculus"
/note="Vector: pY73D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer 15'
TGTACCAACATCTGAAGTGGAGCGGCCGCAATTTTATTTTATTTT 3',
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pY73 vector
(Pharmacia). Library was through one round of
normalization to a Cot - 5. Library constructed by Bento
Soares and M.Fatima Ronaldo. RNA was kindly provided by
Dr. Minoru KO (Wayne State University).
/db_xref="taxon:10090"

```

```

BASE COUNT      94 a      112 c      122 g      91 t
ORIGIN

Query Match      5 6% ; Score 148; DB 34; Length 419;
Best Local Similarity 37.9%; Pred. No. 1,29e-05;
Matches          33; Conservatve 19; Mismatches 35; Indels 0; Gaps 0;

Db              159 CACCTGGGACCAGGTCTTGTCTCGGT 185
                  |  ||'-'|||:|:|  ||:|
CP              412 WAAWWSWTTCNCRAAYTCRTTNARRANGCYTTGNTCKNSRANATRTGNGGYTORA 353
                  |  ||'-'|||:|:|  ||:|

RESULT           8
LOCUS            AA423163                455 bp      mRNA      EST
DEFINITION       ve3db02.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
IMAGE:820203 5' similar to gb:J04953 Mouse getslongin gene, complete
cds (MOSUSe);   mRNA sequence.
AA423163
92101978
AA423163.1 GI:2101978
EST.
house mouse.
Mus musculus.
ORGANISM         Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE        I (bases 1 to 455)
AUTHORS          Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
                 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
                 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
                 Theisinger,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
                 Waterston,R.
TITLE            The WashU-HMI Mouse EST Project
JOURNAL          Unpublished (1996)
COMMENT          On Apr 14, 1993 this sequence version replaced gi:716974.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314.286 1800
Fax: 314.286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:488483
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 165.
Location/Qualifiers
'.1. .455
'/organism="Mus musculus"
'/strain="C57BL/6J"
'/note="Organ: mammary gland; Vector: pRTT3D-Pac
(Pharmacia) with a modified polylinker; Site1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCCGCCGCAGATGTTTTTTTTTTTTTTTTTTT
T 3']"; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pRTT3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
'/db_xref="taxon:10090"
'/clone="IMAGE:820203"
'/clone_lib="Soares mouse mammary gland NbMMG"
```



```
/sex="male"
/issue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"

BASE COUNT      101 a      116 c      134 g      104 t

ORIGIN

Query Match      5.6%; Score 148; DB 13; Length 455;
Best Local Similarity 37.9%; Pred. No. 1.29e-05;
Matches 33; Conservative 19; Mismatches 35; Indels 0; Gaps 0;

Db 50 AGAGGTTCCGGGAGGACTATGCGAGAGACCTGCTACTGATGACGTCTGCTCGGA 109
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 412 WNAHNSWYTCNGCRAAYTCRTTARRAANGCYTNGTNCNSWRAANGRTGNGGTTGRA 353
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 110 CACCTGGACGACGCTGCTGCTGGGT 136
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 352 ANACNGCNACNACCTCTTGTGNGGRT 326
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
LOCUS      W13700      459 bp      mRNA      EST      02-OCT-1997
DEFINITION ma82f05.r1 Soares mouse p3NNF19.5 Mus musculus cDNA clone
IMAGE:317217.5 similar to gb:X04412 GELSOLIN PRECURSOR, PLASMA
(HUMAN);, mRNA sequence.
ACCESSION  W13700
NID         91287738
VERSION     W13700.1 GI:1287738
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus.
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS     Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 459)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On May 8, 1995 this sequence version replaced gi:800256.

TITLE
JOURNAL
COMMENT

Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI: 207833
Seq primer: EMP1rmer
High quality sequence stop: 357.
Location/Qualifiers
1. 459
/organism="Mus musculus"
/Note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCACTGACGAGTGGAGGCGGCGATTTTTTTTTTTT 3'],
TGTACCACTGACGAGTGGAGGCGGCGATTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector.
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Falima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University). "
/db_xref="taxon:10090"
/clone_1ib="Soares mouse p3NNF19.5"
/dev_stage="19.5 dpc total fetus"

BASE COUNT      104 a      123 c      130 g      102 t

ORIGIN

Query Match      5.6%; Score 148; DB 33; Length 459;
Best Local Similarity 37.9%; Pred. No. 1.29e-05;
Matches 33; Conservative 19; Mismatches 35; Indels 0; Gaps 0;

Db 99 AGAGGTTCCGGGAGGACTATGCGAGAGACCTGCTACTGATGACGTCTGCTCGGA 158
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 412 WNAHNSWYTCNGCRAAYTCRTTARRAANGCYTNGTNCNSWRAANGRTGNGGTTGRA 353
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 CACCTGGACGACGCTGCTGCTGGGT 185
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 352 ANACNGCNACNACCTCTTGTGNGGRT 326
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
LOCUS      AA097556      505 bp      mRNA      EST      15-FEB-1997
DEFINITION mo02c08.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone
IMAGE:552398.5 similar to gb:U04953 Mouse gelsolin gene, complete
cds (MOUSE);, mRNA sequence.
ACCESSION  AA097556
NID         91643240
VERSION     AA097556.1 GI:1643240
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus.
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS     Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 505)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1393053.

TITLE
JOURNAL
COMMENT

Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:333190
Seq primer: -28m3 rev1 ET from Amerisham
High quality sequence stop: 356.
Location/Qualifiers
1. 505
/organism="Mus musculus"
/strain="C57BL/6 x CBA"
/Note="Organ: Lung; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. 6-8 month old female lung and 1.5 year old male
lung were source of mRNA. Average insert size: 1.5 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGACGAG
3' -3' adaptor sequence: 5' CTCAGGTTTTTTTTTTTTTTT 3'."
/db_xref="taxon:10090"
/clone_1ib="Stratagene mouse lung 937302"
/clone_1ib="Soares mouse p3NNF19.5"
/dev_stage="6-8 month old"
/lab_host="SOLR (kanamycin resistant)"
/dev_stage="19.5 dpc total fetus"

BASE COUNT      113 a      133 c      147 g      111 t

ORIGIN

Query Match      5.6%; Score 148; DB 36; Length 505;
```



Best Local Similarity 37.9%; Pred. No. 1.29e-05;  
Matches 33; Conservative 19; Mismatches 35; Indels 0; Gaps 0;

Db 100 AGAGTCTCCTGGCGAGCTATGCGAGACGCTAGAGAGCTAGCTGCTCTGGA 159  
CP 412 WNAARWTCNGCGRAATCTTARRAANGCTGNGTCKNSWRAANGTGTGGGTGRA 353  
Db 160 CACCTGGACAGCAGCTTGTCTGGGT 186  
CP 352 ANACNGCNAACNACTCTTGTGTGGGT 326

RESULT 11  
LOCUS AA839460 515 bp mRNA EST 27-FEB-1998  
DEFINITION W50G12.1 Soares mouse mammary gland NBMG Mus musculus cDNA clone  
IMAGE:1247254 5' similar to gb:U04953 Mouse gelsolin gene, complete  
cds (MOUSE);, mRNA sequence.

ACCESSION AA839460  
NID 92915555  
VERSION AA839460.1 GI:2915555  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 515)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.

TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2151013.

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:660942  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 420.

FEATURES  
SOURCE Location/Qualifiers  
1..515

/organism="Mus musculus"  
/strain="C57BL/6J"  
/note="Organ: mammary gland; Vector: pT73D-Pac  
(Pharmacia) with a modified polylinker; Site:1: Not I;  
Site:2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo(dT) primer [5',  
TGTACCAATCTGAAGTGGAGCGCCGCCAATGTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT73 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M.Fatima  
Bonaldo."

/db\_xref="taxon:10090"  
/map="21g"  
/clone="IMAGE:1247254"  
/clone\_lib="Soares mouse mammary gland NBMG"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"

BASE COUNT 113 a 138 c 157 g 107 t  
ORIGIN

Query Match 5.6%; Score 148; DB 18; Length 515;  
Best Local Similarity 37.9%; Pred. No. 1.29e-05;  
Matches 33; Conservative 19; Mismatches 35; Indels 0; Gaps 0;

Db 234 AGAGTCTCCTGGCGAGCTATGCGAGACGCTAGTATGACGTACCTCTGGA 293  
CP 412 WNAARWTCNGCGRAATCTTARRAANGCTGNGTCKNSWRAANGTGTGGGTGRA 353  
Db 294 CACCTGGACAGCAGCTTGTCTGGGT 320  
CP 352 ANACNGCNAACNACTCTTGTGTGGGT 326

RESULT 12  
LOCUS A1019831 522 bp mRNA EST 16-JUN-1998  
DEFINITION u81g06.1 Soares mouse mammary gland NBMG Mus musculus cDNA clone  
IMAGE:1364890 5' similar to gb:U04953 Mouse gelsolin gene, complete  
cds (MOUSE);, mRNA sequence.

ACCESSION A1019831  
NID 93234167  
VERSION A1019831.1 GI:3234167  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 522)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.

TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Jan 17, 1998 this sequence version replaced gi:2044964.

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:898110  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 328.

FEATURES  
SOURCE Location/Qualifiers  
1..522

/organism="Mus musculus"  
/strain="C57BL/6J"  
/note="Organ: mammary gland; Vector: pT73D-Pac  
(Pharmacia) with a modified polylinker; Site:1: Not I;  
Site:2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo(dT) primer [5',  
TGTACCAATCTGAAGTGGAGCGCCGCCAATGTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT73 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M.Fatima  
Bonaldo."

/db\_xref="taxon:10090"  
/clone="IMAGE:1364890"  
/clone\_lib="Soares mouse mammary gland NBMG"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"

BASE COUNT 113 a 133 c 153 g 123 t  
ORIGIN



```
Query Match          5.6% Score 148: DB 14: Length 542:
Best Local Similarity 37.9% Pred. No. 1.23e-05:
Matches 33; Conservative 19; Mismatches 35; Indels 0; Gaps 0;
```



RESULT	15	547 bp	mRNA	EST	12-SEP-1996
LOCUS	W35762				
DEFINITION	mc14d03.r1 Soares mouse		p33MF19.5 Mus musculus	cdna clone	

ACCESSION	W35762
NID	G1317664
VERSION	W35762.1
KEYWORDS	GI:1317664
SOURCE	EST.
ORGANISM	house mouse.
	Mus musculus

REFERENCE 1 (pages 1 to 547)  
AUTHORS Maria, N., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

TITLE The WashU-HHMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On May 8, 1995 this sequence version replaced gi:801502

**CONTACT:** Marra M/Mouse ESF Project  
 Mashu-HHM Mouse ESF Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 850L, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:220261  
 Trace considered overall poor quality  
 Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 1.  
 Location/Qualifiers  
 1..547

```

/organism="Mus musculus"
/ncbi_vector: pRTT3 (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCCATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pRTT3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fátima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University). "
/db_xref="taxon:10090"
/clone="IMAGE:348461"
/clone_lib="Soares mouse p3NMFI9.5"
/dew_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT
125 a 143 c 168 g 111 t
ORIGIN

```

Query Match	5.6%	Score 148;	DB 34;	Length 547;
Best Local Similarity	39.1%	Pred. No. 1,296-05;		
Matches	34;	Conservative	18;	Mismatches 35; Indels 0; Gaps 0;

Db	382	AGACGTCCTCGCAGCGCTATGACGAAAGACCTGCTACTATACGTCACGCTCGTTGA	441
CP	412	WNAWSYTCGCGCAATTCCTTTNRRRANGVYTGTCGTCCKMSPPAANGCTRGNGCYGGA	353

```

Db      442 CTACTGCGACGAGGCTCTTGTGTGAGT 468
      |||||..|||:|:|:|:|:|
Cp      352 ANACNGCMACNACYCTCYTTRTGNGRT 326

```

Search completed: Sat Nov 27 10:18:06 1995  
Job time : 1149 secs;



(77)

Release 3.1A John F. Collins, Biocomputing Research Unit  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

MPSrch_ntp n.a. - n.a. Smith-Waterman search, using a protein database
              which has been backtranslated into n.a. using IUPAC symbols

```

```
Run on:      Wed Nov 24 02:26:20 1999;      MasPar time 7.01 Seconds
Tabular output not generated.                345.981 Million cell updates/sec
```

```

Title: >US-09-103-287-5
Description: (1-19) from US09103287.seq
Perfect Score: 95
N.A. Sequence: 1 CTTCAATTAATGAACGATGC 19
Comp: GAAGTAAATTAAGCTTGCTACG

```

Scoring table: TABLE bkttranslated2  
Gap 40

```
Nmatch      STD :  Dbase 0;  Query 0
```

Searched: 170751 seqs, 63799824 bases x 2

Post-processing: Minimum Match 08

Database:

a-geneq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 29.531; Variance 90.464; scale 0.326

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	76	80.0	204	9	R43567	Human cadherin-10.	5.79e+01
2	76	80.0	241	21	W13133	Partial human cadheri	5.79e+01
3	76	80.0	243	24	W25660	Human cadherin-10.	5.79e+01
4	76	80.0	653	24	W25657	Rat cadherin-10.	5.79e+01
5	76	80.0	653	21	W13128	Full length rat cadhe	5.79e+01
6	71	74.7	46	34	W7686	Staphylococcus aureus	1.70e+02
7	71	74.7	74	34	W62685	Staphylococcus aureus	1.70e+02
8	71	74.7	19	35	W79335	Staphylococcus aureus	1.70e+02
9	71	74.7	27	63	W10344	Maize dwarf mosaic vi	1.70e+02
10	70	73.7	74	39	W89585	Human ATP-binding cas	2.10e+02
11	67	70.5	99	7	R39343	Egib protein.	3.92e+02
12	67	70.5	163	9	R42452	Enzyme involved in ei	3.92e+02

C	13	67	1639.33	M54145	P. falciparum synthet
C	14	67	70.5	S50777	Sequence of the p195
C	15	67	70.5	M37050	S. putrefaciens EPO b
C	16	67	70.5	R99462	Biosynthetic enzyme o
C	17	67	70.5	M55716	H. pylori ORF Olce610
C	18	66	69.5	S55594	H. pylori ORF O9cpg51
C	19	66	69.5	M94358	Rice NAMH-dependent T
C	20	66	69.5	M89964	Antigen from cluster x
C	21	66	69.5	A40615	Sequence of the X-47
C	22	66	69.5	M89881	Protein encoded by cl
C	23	66	69.5	A4121	Partial human cadherin
C	24	66	69.5	R22513	Truncated precursor o
C	25	66	69.5	R88468	Human two chain facto
C	26	66	69.5	A448.7	Factor X.
C	27	66	69.5	R35762	Xylor X (X).
C	28	66	69.5	A452.22	H. pylori protein.
C	29	66	69.5	M40283	Human Factor X protea
C	30	66	69.5	A488.4	Human Factor Xal.
C	31	66	69.5	570.22	Influenza A/Beking/3
C	32	66	69.5	M46713	80 KDa VIPR(a) toxin
C	33	66	69.5	M63794	Bacillus cereus 80 kD
C	34	66	69.5	R91240	B. cereus VIP1 protei
C	35	66	69.5	M68.14	Polypophosphate kinase.
C	36	66	69.5	R77781	Nucleotide sequence o
C	37	66	69.5	M80813	VIPR(a) protein with
C	38	66	69.5	R884.28	B. cereus VIPR(a) pr
C	39	66	69.5	R91239	B. cereus VIPR(a) in
C	40	66	69.5	884.12	Bacillus cereus 100 k
C	41	66	69.5	M46712	100 kDa VIPR(a) toxicl
C	42	66	69.5	1338.17	VIPR2(a)-VIPR(a) pro
C	43	66	69.5	M19520	Maize optimised-B. ce
C	44	66	69.5	M19513	B. cereus VIPR(a)/VI
C	45	66	69.5	3778.3	ACV synthetase.
					4.8ie+02

## ALIGNMENTS

```

RESULT      1
ID          R43567 standard; protein; 204 AA.
AC          R43567;
DE          19-JUN-1994 (first entry)
DT          Human cadherin-10.
KW          Antibodies; binding.
OS          Homo sapiens.
PM          M09323302-A.
PD          28-OCT-1993.
PR          19-APR-1993; U03681.
RA          17-APR-1992; US-872643.
PA          (DOHE-) DOHENY EYE INST.
PI          Suzuki S;
DR          WPI: 93-351714/44.
DT          N-RSDH: 051225.
PT          DNA encoding cadherin 6,7,9 and 10 antibodies to CH 5,6,7,9
FT          and 10 - useful for modulating binding and regulatory activities
PT          of cadherins
PS          Claim 14; Page 75; 99pp; English.
CC          Full length cDNA sequences encoding human homologs of rat cadherins-
CC          4, 8, 11 and 13 and partial cDNAs encoding human homologs of rat
CC          cadherins 6 and 10 were isolated from a human foetal brain cDNA
CC          library using probes obt'd. by amplifying human brain cDNA with
CC          degenerate PCR primers designed from highly conserved segments in
CC          the cytoplasmic domain of mouse N-, E, and P cadherins.
CC          See also R43564-6.
SO          Sequence      204 AA;

Query Match                80.0%; Score 76; DB 9; Length 204;
Best Local Similarity     52.9%; Pred. NO. 5.79e+01;
Matches       9; Conservative 7; Mismatches 1; Indels 0; Gaps 0.

Db         f i n e r
Dt         384 ytytathaygarngny 400
              |||::||::||::| :
Oy         1 CTTCATTATGACCAT 17

```



```

Qt      F I N E R
RESULT  2
ID      W13133 standard; Protein: 241 AA.
AC      W13133:
DE      14-MAY-1997 (first entry)
KW      Partial human cadherin-10.
KW      Ca2+ dependent; cell adhesion protein; foetal; cadherin; rat;
KW      brain; human; antibody; purification; determination;
KW      tissue expression; binding antagonist; calcium ion; partial.
OS      Homo sapiens.
PN      US5597725-A.
PD      28-JAN-1997.
PF      17-APR-1992; 872643.
PR      17-APR-1992; US-872643.
PR      19-APR-1993; US-049460.
PR      26-JAN-1994; US-188228.
PA      (DOHE-) DOHENY EYE INST.
PI      Suzuki S;
DR      WPI; 97-108328/10.
DR      N-PSDB; T61924.
PT      Antibodies to cadherin proteins - useful as cadherin antagonists,
PT      etc.
BS      Example 2: Columns 91-92: 59pp; English.
CC      The present sequence is a partial human cadherin-10, which
CC      is a Ca2+ dependent cell adhesion protein. The human cadherin CDNA
CC      was isolated from a foetal brain CDNA library, using probes based
CC      on homologous rat cadherin CDNA.
CC      Antibodies or fragments that specifically bind the human cadherin
CC      can be used to purify the cadherin, determine its tissue expression
CC      and antagonise its ligand/antiligand binding activities.
SQ      Sequence 241 AA:

Query Match      80.0%; Score 76; DB 21; Length 241;
Best Local Similarity 52.9%; Pred. No. 5.79e+01;
Matches 9; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db      f i n e r
Dt      384 yltlythaaygarngny 400
Oy      :||:||||:||||:|
Qt      1 CTTCAATTATGACGAT 17
        F I N E R

RESULT  3
ID      W25660 standard; Protein: 243 AA.
AC      W25660:
DE      04-NOV-1997 (first entry)
KW      Human cadherin-10.
KW      Human; cadherin; rat; calcium-dependent cell adhesion protein;
KW      superfamily; cytoskeleton; eatenin; cancer.
OS      Homo sapiens.
PN      Key
FH      Location/Qualifiers
FT      Misc_difference 205
FT      /note= "Encoded by 'TA', this residue is omitted in
FT      one reproduction of this sequence given in
FT      the specification"
FT      Misc_difference 229
FT      /note= "Encoded by 'TGA', this residue is omitted in
FT      one reproduction of this sequence given in
FT      the specification"
PN      US5646250-A.
PD      08-JUL-1997.
PF      17-APR-1992; 872643.
PR      19-APR-1993; US-049460.
PR      17-APR-1992; US-872643.
PR      01-NOV-1994; US-332638.
PA      (DOHE-) DOHENY EYE INST.
PI      Suzuki S;
DR      WPI; 97-362997/33.
DR      N-PSDB; T85435.
PT      Human and rat cadherin polypeptide(s) - mediate cell-cell adhesion

```

```

BS      Example 2: Column 83-86; 56pp; English.
CC      This sequence represents human cadherin-10. The invention specifically
CC      provides details of human cadherin-5, -8, -11, -12 and -13, and rat
CC      cadherin-8, -5, -11 and -13. Cadherins are calcium-dependent cell
CC      adhesion proteins. They are glycosylated integral membrane proteins
CC      that have an N-terminal extracellular domain that determines binding
CC      specificity, a hydrophobic membrane spanning region and a C-terminal
CC      cytoplasmic domain, which is highly conserved among members of the
CC      superfamily. The C-terminal domain interacts with the cytoskeleton
CC      through eatenins and other cytoskeleton-associated proteins. The
CC      novel cadherin proteins may be used in the analysis of the role of
CC      cadherins in various cancers. Sequence analysis of the cadherin
CC      proteins also allows investigation of the structure and function of
CC      cadherin. The cadherin proteins may be isolated by using anti-cadherin
CC      antibodies. These antibodies may also be used to modulate the activity
CC      of cadherin and to determine the tissue specific distribution of cadherin
CC      proteins. Each subclass of cadherins has a unique tissue distribution
CC      pattern. The cadherin-10 clone encoding this protein has an open reading
CC      frame which begins at a region corresponding to the middle of the first
CC      extracellular domain of previously identified cadherins.
SQ      Sequence 653 AA:

Query Match      80.0%; Score 76; DB 24; Length 653;
Best Local Similarity 52.9%; Pred. No. 5.79e+01;
Matches 9; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db      f i n e r
Dt      384 yltlythaaygarngny 400
Oy      :||:||||:||||:|
Qt      1 CTTCAATTATGACGAT 17
        F I N E R

RESULT  4
ID      W25657 standard; Protein: 653 AA.
AC      W25657:
DE      04-NOV-1997 (first entry)
KW      Rat cadherin-10;
KW      Human; cadherin; rat; calcium-dependent cell adhesion protein;
KW      superfamily; cytoskeleton; eatenin; cancer; ss.
OS      Rattus rattus.
PN      US5646250-A.
PD      08-JUL-1997.
PF      17-APR-1992; 872643.
PR      19-APR-1993; US-049460.
PR      17-APR-1992; US-872643.
PR      01-NOV-1994; US-332638.
PA      (DOHE-) DOHENY EYE INST.
PI      Suzuki S;
DR      WPI; 97-362997/33.
DR      N-PSDB; T85432.
PT      Human and rat cadherin polypeptide(s) - mediate cell-cell adhesion
PT      Example 1; Column 55-58; 56pp; English.
PS      This sequence represents rat cadherin-10. The invention specifically
CC      provides details of human cadherin-5, -8, -11, -12 and -13, and rat
CC      cadherin-8, -5, -11 and -13. Cadherins are calcium-dependent cell
CC      adhesion proteins. They are glycosylated integral membrane proteins
CC      that have an N-terminal extracellular domain that determines binding
CC      specificity, a hydrophobic membrane spanning region and a C-terminal
CC      cytoplasmic domain, which is highly conserved among members of the
CC      superfamily. The C-terminal domain interacts with the cytoskeleton
CC      through eatenins and other cytoskeleton-associated proteins. The
CC      novel cadherin proteins may be used in the analysis of the role of
CC      cadherins in various cancers. Sequence analysis of the cadherin
CC      proteins also allows investigation of the structure and function of
CC      cadherin. The cadherin proteins may be isolated by using anti-cadherin
CC      antibodies. These antibodies may also be used to modulate the activity
CC      of cadherin and to determine the tissue specific distribution of cadherin
CC      proteins. Each subclass of cadherins has a unique tissue distribution
CC      pattern. The cadherin-10 clone encoding this protein has an open reading
CC      frame which begins at a region corresponding to the middle of the first
CC      extracellular domain of previously identified cadherins.
SQ      Sequence 653 AA:

Query Match      80.0%; Score 76; DB 24; Length 653;

```



Best Local Similarity 52.9%; Pred. No. 5.79e+01;  
Matches 9; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Dt 1731 ytyathaaaygarngy 1747  
Qt 1 CTTCAATATGACGAT 17  
F I N E R

RESULT 5  
ID W13128 standard; Protein; 653 AA.

AC W13128:  
DE 13-MAY-1997 (first entry)  
DE Full length rat cadherin-10.  
KW Ca2+ dependent; cell adhesion protein; foetal; cadherin; rat;  
KW brain; human; antibody; purification; determination;  
KW tissue expression; binding antagonist; calcium ion.  
OS Rattus rattus.

PN US5597725-A.  
PD 28-JAN-1997.  
PE 17-APR-1992; 872643.  
PR 17-APR-1992; US-872643.  
PR 19-APR-1993; US-049460.  
PR 26-JAN-1994; US-188228.  
PA (DOHE-) DOHENY EYE INST.  
PI Suzuki S.  
DR WPI: 97-108328/10.  
DR N-PDB: T61919.

PT Antibodies to cadherin proteins - useful as cadherin antagonists,  
PS etc.

CC Example 1: Columns 61-64; 59pp; English.  
CC The present sequence is full length rat cadherin-10, which is  
CC a Ca2+ dependent cell adhesion protein. The rat cadherin cDNA was  
CC isolated from a brain cDNA library using a labelled rat cadherin-10  
CC extracellular domain PCR fragment as a probe. The rat cDNA was then  
CC used to isolate the cDNA encoding its full length human homologue  
CC from a human foetal brain cDNA library.  
CC Antibodies or fragments that specifically bind the human cadherin  
CC can be used to purify the cadherin, determine its tissue expression  
CC and antagonise its ligand/antiligand binding activities.

SO Sequence 653 AA;

Query Match 80.0%; Score 76; DB 21; Length 653;  
Best Local Similarity 52.9%; Pred. No. 5.79e+01;  
Matches 9; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Dt 1731 ytyathaaaygarngy 1747  
Qt 1 CTTCAATATGACGAT 17  
F I N E R

RESULT 6  
ID W77686 standard; Protein; 46 AA.

AC W77686:  
DE 30-OCT-1998 (first entry)  
DE Staphylococcus aureus protein of unknown function.  
KW Staphylococcus aureus protein; immune response induction; eye infection;  
KW antibody production; T-cell immune response; gastrointestinal infection;  
KW respiratory infection; inhibitor; bacterial infection; cardiac infection;  
KW central nervous system; kidney infection; urinary tract infection;  
KW antimicrobial compound identification; broad spectrum antibiotic;  
KW therapy.

OS Staphylococcus aureus.  
PN EP-841394-A2.  
PD 13-MAY-1998.  
PE 24-SEP-1997; 307485.

PR 24-SEP-1996; US-027032.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PI Black Mt, Burnham MKR, Hodgson JE, Knowles DUC,

PI Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,  
DR WPI: 98-252940/23.  
DR N-PDB: V53479.

PT New nucleic acid sequences from Staphylococcus aureus WCHU29 -  
PT useful in vaccines and for treatment of bacterial infections of e.g.  
PT respiratory tract and central nervous system

PS Claim 11; Page 329; 390pp; English.  
CC This sequence represents a Staphylococcus aureus protein of unknown  
CC function, and is encoded by a DNA sequence of the invention.  
CC The DNA sequences were isolated from Staphylococcus aureus WCHU29  
CC (NCIMB 40771). Host cells containing the DNA sequences are used to  
CC produce polypeptides or fragments. The proteins are used in the treatment  
CC of disease, for inducing an immune response by administering them, to  
CC produce antibody and/or T-cell immune response. Antagonists of the  
CC proteins are used for the inhibition of bacterial polypeptides.  
CC Conditions which may be treated include bacterial infections, especially  
CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,  
CC urinary tract, skin, bones and joints. The proteins can also be used to  
CC identify antimicrobial compounds which are broad spectrum antibiotics,  
CC especially useful in the treatment of H. pylori infection.

Query Match 74.7%; Score 71; DB 34; Length 46;  
Best Local Similarity 50.0%; Pred. No. 1.70e+02;  
Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Dt 4 caymngwnytnatgaar 21  
Ct 18 CAGCGTTCATATGANG 1  
H R S L M K

RESULT 7  
ID W62685 standard; Protein; 74 AA.

AC W62685:  
DE 09-NOV-1998 (first entry)  
DE Streptococcus pneumoniae polypeptide.  
KW Polypeptide; ORF; open reading frame; infection; bacterial;  
KW streptococcal; bacteraemia; diagnosis; prophylaxis.

OS Streptococcus pneumoniae.  
PN W09823631-A1.  
PD 04-JUN-1998.  
PE 24-NOV-1997; U21976.  
PR 27-NOV-1996; US-031879.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PI Black Mt, Hodgson JE, Knowles DUC, Lonetto MA, Nicholas RO,  
PI Reid RH, Zarfos PN;  
DR WPI: 98-322654/28.

PT Streptococcus pneumoniae polynucleotides - useful for developing  
PT products for diagnosis, prevention and treatment of infections e.g.  
PT pneumonia, bacteraemia, meningitis or endocarditis  
PS Claim 5; Page 32; 181pp; English.  
CC The sequence is that of a Streptococcal polypeptide.  
CC The polypeptide can potentially be used for the diagnosis and  
CC prevention of bacterial infections, especially SP infection.  
CC It may be used for the treatment of diseases such as otitis media,  
CC conjunctivitis, pneumonia, bacteraemia, meningitis, sinusitis, pleural  
CC empyema, endocarditis or infection of the cerebrospinal fluid.

Query Match 74.7%; Score 71; DB 34; Length 74;  
Best Local Similarity 47.4%; Pred. No. 1.70e+02;  
Matches 9; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Dt 84 rcaytygwnytnatgaar 102  
Ct 19 CAGCGTTCATATGANG 1  
H R S L M K











CC be produced outside the parasite and has, at least over extended regions,  
 CC the native pattern of folding. Larger amounts of the protein can be  
 CC produced recombinantly than would be possible using the parasites as  
 CC source.

SO Sequence 1639 AA;

Query Match 70.5%; Score 67; DB 33; Length 1639;  
 Best Local Similarity 47.1%; Pred. No. 3.92e+02;  
 Matches 8; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db s l i n e  
 Dc 2046 rwsnytnaahyarg 2062  
 CP 17 ATCGTCATTATGAAG 1  
 Ct S F I N E

RESULT 14

ID P50777 standard; Protein: 1654 AA.

AC P50777;

DT 30-SEP-1991 (first entry)

DE Sequence of the P195 protein of Plasmodium falciparum.

KW Malaria vaccine; epitope; antigen; immunogen.

OS Plasmodium falciparum.

PN EP-154454-A.

PD 11-SEP-1985.

PR 21-FEB-1985; 301173.

PR 22-FEB-1984; GB-004692.

PR 26-SEP-1984; GB-024340.

PR 21-FEB-1985; GB-004429.

PA (WEL) WELLCOME FOUNDATION LTD.

PI Holder A, Sandhu J, Odink K, Lockyer M, Riveros-Moreno V;

DR N-PSDB; N50530.

PT Cloned DNA sequence encoding plasmodium falciparum protein -

PT useful for expressing the protein for use in vaccines against

PT malaria

PS Clam 6; Fig 1; 51pp; English.

CC The sequence encoding the P195 protein of Plasmodium falciparum

CC (N50530) and a peptide comprising at least one of its epitopes

CC (see P50777) are claimed. Also claimed is a vaccine for inducing

CC immunity to malaria comprising the novel peptide or P195 or a

CC peptide comprising at least one epitope when derived from the new

CC DNA sequence, together with a carrier.

SO Sequence 1654 AA;

Query Match 70.5%; Score 67; DB 3; Length 1654;

Best Local Similarity 47.1%; Pred. No. 3.92e+02;

Matches 8; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db s l i n e  
 Dc 2046 rwsnytnaahyarg 2062  
 CP 17 ATCGTCATTATGAAG 1  
 Ct S F I N E

RESULT 15

ID M37050 standard; Protein: 2756 AA.

AC M37050;

DT 03-JUL-1998 (first entry)

DE S. putrefaciens EPO biosynthesis gene cluster ORF6 product.

KW SCRC-2874; FERM BP-1625; eicosapentaenoic acid; EPA;

KW biosynthesis gene cluster; synthetase.

OS Shewanella putrefaciens.

PN W09801565-A1.

PD 15-JAN-1998.

PR 09-JUL-1997; J02371.

PR 10-JUL-1996; JP-180845.

PA (SAGA) SAGAMI CHEM RES CENTRE.

PI Kato S, Rondo K, Yamada A, Yazawa K;

DR WPI: 98-101060/09.

N-PSDB; V00503.

PT Eicosapentaenoic acid produced by culture of transformed Escherichia  
 PT coli - containing an eicosapentaenoic acid synthetase gene derived  
 PT from the marine microorganism Shewanella  
 PS Example 1; Pages 53-73; 110pp; Japanese.

CC The present sequence is encoded by the Shewanella putrefaciens

CC SCRC-2874 (FERM BP-1625) eicosapentaenoic acid (EPA) biosynthesis

CC gene cluster.

CC A novel EPA (useful in drugs, pesticides, foods and feedstuffs) is

CC encoded by synthetase enzyme gene sequences comprising parts of the

CC full sequence of the synthetase gene from the marine microorganism

CC S. putrefaciens SCRC-2874 (FERM BP-1625), in which at least 1 of

CC the 9 open reading frames (ORF) (numbered 2-10) in the gene have

CC been deleted. In particular the gene sequences comprising the

CC following parts of the full gene:

CC (1) bases 8081-9441, 12314-13084 and 13889-32520;

CC (2) bases 8081-9441, 12314-13084, 13889-32520 and 34627-35559;

CC (3) bases 8081-9441, 12314-13084 and 13889-35559;

CC (4) bases 8081-9441, 9681-13084 and 13889-32520;

CC (5) bases 8081-9441, 9681-13084, 13889-32520 and 34627-35564; and

CC (6) bases 8081-9441, 9681-13084 and 13889-35564, are claimed.

SO Sequence 2756 AA;

Query Match 70.5%; Score 67; DB 30; Length 2756;

Best Local Similarity 50.0%; Pred. No. 3.92e+02;

Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db a s l i n e  
 Dc 1549 gcnwsnytnaahyarg 1566  
 CP 19 GCATCGTCATTATGAAG 2  
 Ct A S F I N E

Search completed: Wed Nov 24 02:28:25 1999  
 Job time : 125 secs.



\*\*\*\*\*  
MIRAGE (TM)  
\*\*\*\*\*

Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPerch\_nrp n.a. - n.a. Smith-Waterman search, using a protein database  
which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Nov 24 02:29:02 1999; MasPar time 3.03 Seconds

Tabular output not generated. 448,944 Million cell updates/sec

Title: >US-09-103-287-5  
Description: (1-19) from US09103287.seq  
Perfect Score: 95  
N.A. Sequence: 1 CTTCATTATGACGATGC 19  
Comp: GAAGTATATCTGCTACG

Scoring table: TABLE bkttranslated  
Gap 40

Mmatch STD : Dbase 0; Query 0

Searched: 122461 seqs, 35738955 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCR9\_COMB 4:backfiles1

Statistics: Mean 28.081; Variance 92.387; scale 0.304

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	76	80.0	241	1	US-08-188-Sequence 56, Applicati	3.90e+01
2	76	80.0	241	1	US-08-332-Sequence 56, Applicati	3.90e+01
3	76	80.0	241	1	US-08-332-Sequence 50, Applicati	3.90e+01
4	76	80.0	653	1	US-08-188-Sequence 46, Applicati	3.90e+01
5	76	80.0	653	1	US-08-332-Sequence 46, Applicati	3.90e+01
6	70	73.7	747	2	US-08-895-Sequence 1, Applicatio	1.35e+02
7	67	70.5	990	2	US-08-392-Sequence 20, Applicati	2.46e+02
8	67	70.5	990	2	US-08-466-Sequence 15, Applicati	2.46e+02
9	67	70.5	2756	2	US-08-752-Sequence 11, Applicati	2.46e+02
10	67	70.5	2756	1	US-08-375-Sequence 11, Applicati	2.46e+02
11	67	70.5	2756	1	US-08-933-Sequence 123, Applicat	3.00e+02
12	66	69.5	17	2	US-08-934-Sequence 472, Applicat	3.00e+02
13	66	69.5	26	1	US-07-942-Sequence 13, Applicati	3.00e+02
14	66	69.5	224	2	US-08-766-Sequence 4, Applicatio	3.00e+02
15	66	69.5	241	1	US-08-474-Sequence 4, Applicatio	3.00e+02
16	66	69.5	241	1	US-08-774-Sequence 4, Applicatio	3.00e+02
17	66	69.5	241	1	US-08-484-Sequence 4, Applicatio	3.00e+02
18	66	69.5	254	1	US-08-474-Sequence 3, Applicatio	3.00e+02
19	66	69.5	254	1	US-08-474-Sequence 3, Applicatio	3.00e+02

RESULT	ID	Sequence	Standard	PRT	241 AA
20	66	69.5	254	1	US-08-330-Sequence 3, Applicatio
21	66	69.5	306	1	US-08-330-Sequence 1, Applicatio
22	66	69.5	306	1	US-08-484-Sequence 1, Applicatio
23	66	69.5	306	1	US-08-474-Sequence 1, Applicatio
24	66	69.5	570	2	US-08-453-Sequence 7, Applicatio
25	66	69.5	571	2	US-08-453-Sequence 15, Applicati
26	66	69.5	667	2	US-08-471-Sequence 7, Applicatio
27	66	69.5	667	2	US-08-471-Sequence 7, Applicatio
28	66	69.5	667	2	US-08-471-Sequence 7, Applicatio
29	66	69.5	852	2	US-08-471-Sequence 36, Applicati
30	66	69.5	852	2	US-08-470-Sequence 36, Applicati
31	66	69.5	852	2	US-08-471-Sequence 36, Applicati
32	66	69.5	884	2	US-08-470-Sequence 5, Applicatio
33	66	69.5	884	2	US-08-471-Sequence 5, Applicatio
34	66	69.5	884	2	US-08-471-Sequence 5, Applicatio
35	66	69.5	900	2	US-09-005-Sequence 62, Applicati
36	66	69.5	900	2	US-08-630-Sequence 62, Applicati
37	66	69.5	913	2	US-08-472-Sequence 5, Applicatio
38	66	69.5	913	2	US-08-474-Sequence 6, Applicatio
39	66	69.5	913	2	US-08-474-Sequence 6, Applicatio
40	66	69.5	916	2	US-08-188-Sequence 48, Applicati
41	66	69.5	1338	2	US-08-471-Sequence 50, Applicati
42	66	69.5	1338	2	US-08-471-Sequence 50, Applicati
43	66	69.5	1338	2	US-08-463-Sequence 50, Applicati
44	66	69.5	1346	2	US-08-471-Sequence 23, Applicati
45	66	69.5	1346	2	US-08-471-Sequence 23, Applicati

## ALIGNMENTS

Sequence 56, Application US/08188228  
Patent No. 5597725  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray &  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/188,228  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/049,460  
FILING DATE: 19 APR 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/872,643  
FILING DATE: 17 APR 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5597725and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31340  
TELECOMMUNICATION INFORMATION:



CC TELEPHONE: (312) 474-6300  
CC TELEFAX: (312) 474-0448  
CC TELE: 25-3856  
CC INFORMATION FOR SEQ ID NO: 56:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 241 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 241 AA; 26701 MW; 300025 CN;

Query Match 80.0%; Score 76; DB 1; Length 241;  
Best Local Similarity 52.9%; Pred. No. 3.90e+01;  
Matches 9; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

DB 384 YTTYATHAAGMGNV 400  
DT :||:||||:||||:|  
DY 1 CTTCAATTAATGACGAT 17  
QT F I N E R

RESULT 2  
ID US-08-332-638-56 STANDARD; PRT: 241 AA.  
XX xxxxxx

Sequence 56, Application US/08332638  
CC Patent No. 5646250  
CC GENERAL INFORMATION:  
CC APPLICANT: Suzuki, Shintaro  
CC TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS  
CC NUMBER OF SEQUENCES: 62  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
CC STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CC CITY: Chicago  
CC STATE: Illinois  
CC COUNTRY: USA  
CC ZIP: 60606  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/332,638  
CC FILING DATE: 01-NOV-1994  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/872,643  
CC FILING DATE: 17 APR 1992  
CC APPLICATION NUMBER: US/08/049,460  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: No. 5646250and, Greta E.  
CC REGISTRATION NUMBER: 35,302  
CC REFERENCE/DOCKET NUMBER: 31340  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (312) 474-6300  
CC TELEFAX: (312) 474-0448  
CC TELE: 25-3856  
CC INFORMATION FOR SEQ ID NO: 56:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 241 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein

SEQ SEQUENCE 241 AA; 26701 MW; 300025 CN;

Query Match 80.0%; Score 76; DB 1; Length 241;  
Best Local Similarity 52.9%; Pred. No. 3.90e+01;  
Matches 9; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

DB 384 YTTYATHAAGMGNV 400  
DT :||:||||:||||:|  
DY 1 CTTCAATTAATGACGAT 17  
QT F I N E R

RESULT 3  
ID US-08-332-643-50 STANDARD; PRT: 241 AA.  
XX xxxxxx

Sequence 50, Application US/08332643  
CC Patent No. 5639634  
CC GENERAL INFORMATION:  
CC APPLICANT: Suzuki, Shintaro  
CC TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS  
CC NUMBER OF SEQUENCES: 56  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
CC STREET: Two First National Plaza, 20 South Clark  
CC CITY: Chicago  
CC STATE: Illinois  
CC COUNTRY: USA  
CC ZIP: 60603  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/332,643  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/872,643  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: No. 5639634and, Greta E.  
CC REGISTRATION NUMBER: 35,302  
CC REFERENCE/DOCKET NUMBER: 27866/30795  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (312) 346-5750  
CC TELEFAX: (312) 984-9740  
CC TELE: 25-3856  
CC INFORMATION FOR SEQ ID NO: 50:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 241 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 241 AA; 26701 MW; 300025 CN;

Query Match 80.0%; Score 76; DB 1; Length 241;  
Best Local Similarity 52.9%; Pred. No. 3.90e+01;  
Matches 9; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

DB 384 YTTYATHAAGMGNV 400  
DT :||:||||:||||:|  
DY 1 CTTCAATTAATGACGAT 17  
QT F I N E R



Qt F I N E R

RESULT 4  
ID US-08-188-228-46 STANDARD: PRT: 653 AA.

XX xxxxxx

DE Sequence 46, Application US/08188228

CC Sequence 46, Application US/08188228  
CC Patent No. 5597725

CC GENERAL INFORMATION:

CC APPLICANT: Suzuki, Shintaro

CC TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS

CC NUMBER OF SEQUENCES: 62

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &amp;

CC ADDRESSEE: Borun

CC STREET: 6300 Sears Tower, 233 S. Wacker Drive

CC CITY: Chicago

CC STATE: Illinois

CC COUNTRY: USA

CC ZIP: 60606

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patentin Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC ATTORNEY/AGENT INFORMATION:

CC FILING DATE: 17 APR 1992

CC REGISTRATION NUMBER: 35,302

CC REFERENCE/DOCKET NUMBER: 31340

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (312) 474-6300

CC TELEFAX: (312) 474-0448

CC TELEX: 25-3856

CC INFORMATION FOR SEQ ID NO: 46:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 653 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 653 AA; 72459 MW; 2123017 CN;

SQ

Query Match 80.0%; Score 76; DB 1; Length 653;

Best Local Similarity 52.9%; Pred. No. 3.90e+01;

Matches 9; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db F I N E R

Dt 1731 YTTYATHAAYGARMGNV 1747

Dy 1 CTTCATTAATGACGAT 17

Qt F I N E R

RESULT 5  
ID US-08-332-638-46 STANDARD: PRT: 653 AA.

XX xxxxxx

Dt Sequence 46, Application US/08332638

DE Sequence 46, Application US/08332638

CC Patent No. 5646250

CC GENERAL INFORMATION:

CC APPLICANT: Suzuki, Shintaro

CC TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS

CC NUMBER OF SEQUENCES: 62

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &amp;

CC ADDRESSEE: Borun

CC STREET: 6300 Sears Tower, 233 S. Wacker Drive

CC CITY: Chicago

CC STATE: Illinois

CC COUNTRY: USA

CC ZIP: 60606

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patentin Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/332,638

CC FILING DATE: 01-NOV-1994

CC CLASSIFICATION: 435

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 07/872,643

CC FILING DATE: 17 APR 1992

CC ATTORNEY/AGENT INFORMATION:

CC NAME: No. 5646250and, Greta E.

CC REGISTRATION NUMBER: 35,302

CC REFERENCE/DOCKET NUMBER: 31340

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (312) 474-6300

CC TELEFAX: (312) 474-0448

CC TELEX: 25-3856

CC INFORMATION FOR SEQ ID NO: 46:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 653 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 653 AA; 72459 MW; 2123017 CN;

SQ

Query Match 80.0%; Score 76; DB 1; Length 653;

Best Local Similarity 52.9%; Pred. No. 3.90e+01;

Matches 9; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db F I N E R

Dt 1731 YTTYATHAAYGARMGNV 1747

Dy 1 CTTCATTAATGACGAT 17

Qt F I N E R

RESULT 6  
ID US-08-895-522-1 STANDARD: PRT: 747 AA.

XX xxxxxx

DE Sequence 1, Application US/08895522

CC Patent No. 5858719

CC GENERAL INFORMATION:

CC APPLICANT: Hillman, Jennifer L.

CC APPLICANT: Shah, Puri



CC APPLICANT: Corley, Neil C.  
CC TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE  
CC TITLE OF INVENTION: TRANSPORT PROTEIN  
CC NUMBER OF SEQUENCES: 4  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Incyte Pharmaceuticals, Inc.  
CC STREET: 3174 Porter Dr.  
CC CITY: Palo Alto  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 94304  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette  
CC COMPUTER: IBM Compatible  
CC OPERATING SYSTEM: DOS  
CC SOFTWARE: FASTSEQ for Windows Version 2.0  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/895,522  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER:  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Billings, Lucy J  
CC REGISTRATION NUMBER: 36,749  
CC REFERENCE/DOCKET NUMBER: PF-0336 US  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 415-855-0555  
CC TELEFAX: 415-845-4166  
CC TELEX:  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 747 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC IMMEDIATE SOURCE:  
CC LIBRARY: OVARNOT02  
CC CLONE: 545981  
SQ SEQUENCE 747 AA; 82138 MW; 2890233 CN;  
Db F N N E R  
Dt 1020 YTTAAATGAGAGAGT 1036  
Qt 1 CTTCAATGACGAT 17  
F I N E R  
Query Match 73.7%; Score 70; DB 2; Length 747;  
Best Local Similarity 52.9%; Pred. No. 1.35e+02;  
Matches 9; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

CC APPLICANT: Wieland, Bernd  
CC APPLICANT: Kupke, Thomas  
CC APPLICANT: Jung, G nther  
CC APPLICANT: Kellner, Roland  
CC TITLE OF INVENTION: Biosynthetic Process for The Preparation  
CC OF SEQUENCES: 42  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
CC STREET: 1100 New York Avenue  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: U.S.A.  
CC ZIP: 20005  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/392,625  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/876,791  
CC FILING DATE: 30-APR-1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Esmond, Robert W.  
CC REGISTRATION NUMBER: 32,893  
CC REFERENCE/DOCKET NUMBER: 0652.0980002  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (202) 371-2600  
CC TELEFAX: (202) 371-2540  
CC INFORMATION FOR SEQ ID NO: 20:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 990 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
SQ SEQUENCE 990 AA; 117141 MW; 5092259 CN;  
Db S F I N E  
Dt 2046 RWSYTAATGAGAGT 2061  
Ct 17 ATCCTCATATGAA 2  
S F I N E  
Query Match 70.5%; Score 67; DB 2; Length 990;  
Best Local Similarity 50.0%; Pred. No. 2.46e+02;  
Matches 8; Conservative 7; Mismatches 1; Indels 0; Gaps 0;











```

CC      NAME: Isaacson, John P.
CC      REGISTRATION NUMBER: 33,751
CC      REFERENCE/DOCKET NUMBER: 040433/0148
CC      INFORMATION FOR SEQ ID NO: 123:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 17 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      SEQUENCE 17 AA; 1885 MW; 1336 CN;
SQ
Query Match          69.5%; Score 66; DB 2; Length 17;
Best Local Similarity 52.6%; Pred. No. 3,00e+02;
Matches 10; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
Db      A L L I N E
Dt      13 GCATTNTNATHAAYGARG 31
        || : |||:|:|:|:|
Cc      19 GCATCGTTCATTATGAAG 1
        A S F I N E
Cc
RESULT 14
ID      US-07-942-245-472      STANDARD;      PRT: 26 AA.
XX      AC      xxxxxx
XX      DT
XX
DE      Sequence 472, Application US/07942245
CC      Sequence 472, Application US/07942245
CC      Patent No. 5639641
CC      GENERAL INFORMATION:
CC      APPLICANT: PEDERSEN, Jan T.
CC      APPLICANT: SEARLE, Stephen M.J.
CC      APPLICANT: REES, Anthony R.
CC      APPLICANT: ROGUSKA, Michael A.
CC      APPLICANT: GUILD, Braydon C.
CC      TITLE OF INVENTION: SURFACE RESIDUE VENERERING OF RODENT
CC      TITLE OF INVENTION: ANTIBODIES
CC      NUMBER OF SEQUENCES: 522
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
CC      STREET: 2100 Pennsylvania Avenue, N.W.
CC      CITY: Washington
CC      STATE: D.C.
CC      COUNTRY: United States
CC      ZIP: 20037-3202
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: HP 9000/700 workstation
CC      OPERATING SYSTEM: UNIX
CC      SOFTWARE: In house
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/07/942,245
CC      FILING DATE: 09-SEP-1992
CC      CLASSIFICATION: 530
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (202) 293-7060
CC      TELEFAX: (202) 293-7860
CC      TELEX: 6491103
CC      INFORMATION FOR SEQ ID NO: 472:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 26 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: Peptide
CC      SEQUENCE 26 AA; 2950 MW; 3814 CN;
SQ
Query Match          69.5%; Score 66; DB 1; Length 26;
Best Local Similarity 50.0%; Pred. No. 3,00e+02;
Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

```







\*\*\*\*\*  
 M O S E R  
 (TM)  
 \*\*\*\*\*

Release 3.1a John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Mpserch\_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 12:37:31 1999; MasPar time 90.60 Seconds

Tabular output not generated. 581.322 Million cell updates/sec

Title: >US-09-103-287-5

Description: (1-19) from US09103287.seq

Perfect Score: 19

N.A. Sequence: 1 CTTTCATTATGACGATGC 19

Comp: GAAGTATATGCTGCTACG

Scoring table: TABLE default

Gap 10

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%

Database: Listing first 45 summaries

Database:

emb158  
 1:em\_ba1 2:em\_ba2 3:em\_fun 4:em\_hun1 5:em\_hun2  
 7:em\_in 8:em\_com 9:em\_or 10:em\_h11 11:em\_pat 12:em\_ph  
 13:em\_p1 14:em\_ro 15:em\_sts 16:em\_v1  
 genbank111  
 17:gb\_pa1 18:gb\_pa2 19:gb\_hgt1 20:gb\_hgt2 21:gb\_in1  
 22:gb\_in2 23:gb\_com 24:gb\_ov 25:gb\_pat 26:gb\_ph 27:gb\_p11  
 28:gb\_p12 29:gb\_p1 30:gb\_p12 31:gb\_p13 32:gb\_ro  
 33:gb\_st 34:gb\_sts 35:gb\_sy 36:gb\_un 37:gb\_v1

Statistics: Mean 6.442; Variance 3.042; scale 2.118

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	ID	Description	Pred. No.
C 1	89.5	1314	17	AF034076	8.68e+00
C 2	84.2	3762	17	B0015178	3.46e+01
C 3	84.2	279110	17	RPMX03	1.32e+02
C 4	78.9	1489	27	SPPTGIVAB	1.32e+02
C 5	78.9	1489	27	SPPTGIVAB	1.32e+02
C 6	78.9	2303	27	KSPGTIVAC	1.32e+02
C 7	78.9	2638	21	EHU04863	1.32e+02
C 8	78.9	2652	21	EHNADDAD	1.32e+02
C 9	78.9	3044	27	SPCEN2B	1.32e+02
C 10	78.9	3866	18	ANANIFJ	1.32e+02
C 11	78.9	3878	28	TAU51330	1.32e+02
C 12	78.9	3919	27	YSCRAD2G	1.32e+02
C 13	78.9	5175	27	SCYGR238C	1.32e+02

#### ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	NID	VERSION	KEYWORDS	SOURCE	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
C 14	15	78.9	5565	18	AF005098	Lactococcus lactis RNA	1.32e+02					
C 15	15	78.9	5791	24	CHKVITC	Chicken vitellogenin g	1.32e+02					
C 16	15	78.9	10534	27	SCCHRVII	S.cerevisiae genomic s	1.32e+02					
C 17	15	78.9	11093	18	US32799	Haemophilus influenzae	1.32e+02					
C 18	15	78.9	20343	24	GGVITIG	Chicken vitellogenin I	1.32e+02					
C 19	15	78.9	28735	21	CEIC02D5	C. elegans cosmid C02D	1.32e+02					
C 20	15	78.9	39489	30	HSU69570	Human xpd2 cosmid U27H	1.32e+02					
C 21	15	78.9	41351	21	CELF42C5	Caenorhabditis elegans	1.32e+02					
C 22	15	78.9	76169	30	AC003014	Human PAC clone DJ290B	1.32e+02					
C 23	15	78.9	113546	19	HSAC000382	Human chromosome 11 pO	1.32e+02					
C 24	15	78.9	132340	20	AC002518	Human chromosome X, WO	1.32e+02					
C 25	15	78.9	147686	31	AC005702	Homo sapiens clone GSI	1.32e+02					
C 26	15	78.9	147686	31	AC005702	Homo sapiens chromosome	1.32e+02					
C 27	15	78.9	199517	19	AC006755	Caenorhabditis elegans	1.32e+02					
C 28	14	73.7	436	34	G47706	zebrafish AB	4.77e+02					
C 29	14	73.7	2204	27	D63342	Bacteriophage P4 gop,	4.77e+02					
C 30	14	73.7	2640	26	PP4G0PC	Bacteriophage P4 gop,	4.77e+02					
C 31	14	73.7	2838	17	BCO16GV	Bacillus cereus gene f	4.77e+02					
C 32	14	73.7	4864	22	NA097150	Nosema apis small subu	4.77e+02					
C 33	14	73.7	5910	17	AB001896	Staphylococcus aureus	4.77e+02					
C 34	14	73.7	11147	18	MPAE000018	Mycoplasma pneumoniae	4.77e+02					
C 35	14	73.7	31205	21	CEIC06G1	Caenorhabditis elegans	4.77e+02					
C 36	14	73.7	36493	27	SPBC262	S.pombe chromosome II	4.77e+02					
C 37	14	73.7	42000	29	AC005929	Leishmania major chrom	4.77e+02					
C 38	14	73.7	109821	29	HS10C16	Human DNA sequence fro	4.77e+02					
C 39	14	73.7	117101	30	AC004061	Homo sapiens chromosome	4.77e+02					
C 40	14	73.7	120376	29	HS271G9	Human DNA sequence fro	4.77e+02					
C 41	14	73.7	139990	29	AC006314	Homo sapiens clone DJ0	4.77e+02					
C 42	14	73.7	140709	29	HS103M2	Human DNA sequence fro	4.77e+02					
C 43	14	73.7	150613	27	AB001684	Chlorella vulgaris C-2	4.77e+02					
C 44	14	73.7	180551	30	HUAC004158	Homo sapiens Chromosom	4.77e+02					
C 45	14	73.7	330740	37	PBU42580	Paramecium bursaria Ch	4.77e+02					

AF034076 1314 bp DNA BCT 26-NOV-1997  
 Staphylococcus aureus UDP-N-acetylmuramoyl-L-alanine synthetase  
 AF034076  
 92642658  
 AF034076.1 GI:2642658  
 Staphylococcus aureus.  
 Staphylococcus aureus  
 Bacteria; Firmicutes; Low G+C gram-positive bacteria;  
 Bacillaceae; Staphylococcus.  
 1 (bases 1 to 1314)  
 Lowe, A.M. and Deresiewicz, R.L.  
 Cloning and sequencing of Staphylococcus aureus murC, a gene  
 involved in cell wall biosynthesis  
 Unpublished  
 2 (bases 1 to 1314)  
 Lowe, A.M. and Deresiewicz, R.L.  
 Direct Submission  
 Submitted (11-NOV-1997) Channing Laboratory, Brigham and Women's  
 Hospital and Harvard Medical School, 181 Longwood Ave, Boston, MA  
 02115, USA  
 Location/Qualifiers  
 1..1314  
 /organism="Staphylococcus aureus"  
 /db\_xref="taxon:1280"  
 1..1314  
 /gene="murC"  
 1..1314  
 /gene="murC"  
 /function="cell wall biosynthesis"  
 /note="murC: UDP-N-acetylmuramate-alanine ligase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="UDP-N-acetylmuramoyl-L-alanine synthetase"  
 /protein\_id="AAB87090.1"







```
/protein_id="CA14935.1"
/db_xref="PID:e1342779"
/db_xref="PID:g3861035"
/db_xref="GI:3861035"
/translation="MRHRIINLRKLFIEYNIEGYIIPSDRKNMEYVEAKRLEYI
TGFSSGAIICDAAPEFTGRYLENNKLEDAFYKIDOLKEIYTLIDKNIGY
DQULTYOVLANTLNINFKINENLVKINRPLEPNSVYIADIKFAGVSNDRK
CRETILSSSVTGCNKNDLVLIDSSICWLNLRSDVNTPLFAKYLITSTK
LYLFINPRIIDEIINAREPITLPEKEENILRDSKNRYLIDDSITVHMDLANK
KVKKIVEPCLAKACKNDIEIKHAIDFHIDKVALCEFAEFELHSSSENVCSFH
EITHESLCLKLTAORAKOEGVDSFHAICFOENSATIHYPANPKTAKLEGSHL
LIDSGAOKGATDTITRTITVIGIPCEOKKRYTOYKGIATLRAKPEPNITVGT
ILAROYLMDMIDYRHGTGHGVSFLSVHESQSTINLSKTIILKAGMILSNPGYIP
GTGRIENLITVKNNGWLEFETLSVPAASKLIDVALNIDEINYLKEIYKILRAK
IYNLSTOKRYLWNEINLFQSL"
/gene="RP483"
/complement(3174..4463)
/gene="RP483"
/codon_start=1
/transl_table=11
/product="PUTRESCINE-ORNITHINE ANTIporter (pote)"
/protein_id="CA14936.1"
/db_xref="PID:e1342780"
/db_xref="PID:g3861036"
/db_xref="GI:3861036"
/translation="MSQKLSFNAVFALVTGSGTGVFLPLSLAPFGVYSIMGMLS
LEGMSIALVESCICTKPKTGPHYVASGDIAPFTGWTYIISFVSIVIS
AIGYLPKFSQTLIDLIQLILALAILNKGPIAGKVEYFLTLKFLVLVGL
ALFHNIDNIVIAKEVENETIPTIMGRALLPFGECATTTAGVKNPATPI
RAIITGCVAFIYIINSIGITGLIPASLINSKAPYADATILFEGTSKATYIAS
VICITLNAWVLISGQIALGLAEDGLPFEKKNNAFTGIIISCGITPLLEF
SNNNAKQITQIIDSVAIFLVYLICSLAFLKVSSENSYYLLFVAIISILECT
WVIYKTFETLTIACSFITLIGIPYVGYWRLMNRKL"
/complement(4633..4965)
/gene="RP484"
/complement(4633..4965)
/gene="RP484"
/codon_start=1
/transl_table=11
/product="HESB PROTEIN (hesb)"
/protein_id="CA14937.1"
/db_xref="PID:e1342781"
/db_xref="PID:g3861037"
/db_xref="GI:3861037"
/translation="MKNVILSIDAAKQVLLIERAKPTFGIRVIGSGCAGOTY
VEYADNKNQFDEVVEKGVRLIDPKTLMYILGSEMDYETNFKSQFTPTNNEKANC
GCGKSPSV"
/complement(4969..5364)
/gene="RP485"
/complement(4969..5364)
/gene="RP485"
/codon_start=1
/transl_table=11
/product="NIFU PROTEIN (nifu)"
/protein_id="CA14938.1"
/db_xref="PID:e1342782"
/db_xref="PID:g3861038"
/db_xref="GI:3861038"
/translation="MAYSKVVDHYENPRNNGSLDKKKNVGTGLVAPACDVAKLO
LEVGDDEITDAKFTFGGSAIASSSTETIKGSVEDAKEIKNTETAKELSLPVP
KLHCSLADDAIKAAIADYKOKRENNKDS"
/complement(5430..6662)
/gene="RP486"
/complement(5430..6662)
/gene="RP486"
/codon_start=1
/transl_table=11
/product="NIFS PROTEIN HOMOLOG (spl1)"
/protein_id="CA14939.1"
/db_xref="PID:e1342783"
/db_xref="PID:g3861039"
/db_xref="GI:3861039"
```

```
/translation="MNQKMLTPIYMDVOSTPIPRVMEAMLPYETTKFQNPDR
SHSFEWENAVENARSKAVYIGADSEIIFTSGATSSNNLVYKIGAKFTGNKKH
ITLVSEKCVLNACRHLQEGIKITYLEIKSNGIIDLLETLNALTQTLVAVAVN
EIGVQPLKEIKICRERNVFEPSDIAQFCKIPINVECNIDLASISGHIYGPKI
GALYIRKPRVRYRVPLELVNGOERMSRGTPTPLVGLGASLSEIAYEMKQDHVN
YLDFEILNHSKISLEYVINGDKOORVGNINISPAVGEISIIAIDVLSSGAC
TSASLEPSYVARSISISELHNTSIRFQIGFTTQOEIDYAVNLVCSIDKRLRSL
WEMQEGVDLKIEMTAH"
/complement(6740..7870)
/gene="RP487"
/complement(6740..7870)
/gene="RP487"
/codon_start=1
/transl_table=11
/product="NIFS PROTEIN HOMOLOG (spl1)"
/protein_id="CA14940.1"
/db_xref="PID:e1342784"
/db_xref="PID:g3861040"
/db_xref="GI:3861040"
/translation="MAYLDHNAFTFIDPRVKEYIISLMDKELNPSSAHTSGRAKNI
ETASQIATSIGITWISREYNITFTSSGESNNLIMKNYDGDIFISALEHLSYHIN
YAPNKKIRVKNQGLVDEHLEDLSGNASKLVSVMANNENGVLDQIAISKTK
KYNAFHSDIVOSFEGKISLNTKEIGLDPATISGHIIGGCGGALISNQLPILII
GGCGKSVRSCTENVLAIFGIAAEPRKDISKRYIKIRYLOEELBOKLEYSVNI
ISNNVARLNTTLIVNTDQKVLIGDLNLCVSSGACSSGKISKSHVLTNMGIR
EEARSSIRISLSHNTISDIKAFKAEELYEKSSCNHC1"
/complement(7931..8125)
/gene="RP488"
/complement(7931..8125)
/gene="RP488"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA14941.1"
/db_xref="PID:e1342785"
/db_xref="PID:g3861041"
/db_xref="GI:3861041"
/translation="WILKIKETVYMAILELATNINSGPIILDKYSIKHNSLNYLAQ
IFSKLKSDDLKVSIIIESKDC"
/complement(8277..9143)
/gene="RP489"
/complement(8277..9143)
/gene="RP489"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA14942.1"
/db_xref="PID:e1342786"
/db_xref="PID:g3861042"
/db_xref="GI:3861042"
/translation="MYLISFLIGLNTLRLRLSISSIDEYKDYKKNYGYGIKYLE
TLSEFPIIYCFILNITLTKDYNGSSKVDNINYEIINOLPKIKYNSKISVEE
VEPIYLYSKNNKIFVIDTKNOVSOEKSVPYVEEKKLINVSNRKKHPSVNY
SEIFKONVILPELIKRYPADNLEIRANLNFCTPTVITLFWVTFILLESIVLAV
YGLANLTYTKTSIQTSIRLVWSSGIPILIDPVIIILLPELSILLQLLMFTFLVAV
AILQINKSLSH"
/complement(9681..10349)
/gene="RP490"
/complement(9681..10349)
/gene="RP490"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA14943.1"
/db_xref="PID:e1342787"
```

Note: remainder of annotations omitted.

Query Match 84.2%; Score 16; DB 17; Length 279110;  
Best local similarity 94.4%; Pred. No. 3,46+01;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
DB 101063 GCATCGTTATTAATGAA 101080



CP 19 GCATCCTTCAATGAG 2

```

RESULT 4 YSPGTGIVAB 1489 bp DNA PLN 28-APR-1995
LOCUS S.pombe Ile-1,Ala-1 and Val-1 tRNA genes.
DEFINITION M57699
ACCESSION 9173497
MID 9173497
VERSION M57699.1 GI:173497
KEYWORDS transfer RNA; transfer RNA-Ala; transfer RNA-Ile-I; transfer RNA-Val-I.
SOURCE Schizosaccharomyces pombe DNA.
ORGANISM Schizosaccharomyces pombe
Eukaryotae; mitochondria; eukaryotes; Fungi; Ascomycota;
Archaeoscomycetes; Schizosaccharomycetales;
Schizosaccharomycetaceae; Schizosaccharomyces.
REFERENCE 1 (bases 1 to 1489)
AUTHORS Kuhn,R.M., Clarke,L. and Carbon,J.
TITLE Clustered tRNA genes in Schizosaccharomyces pombe centromeric DNA
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88 (4), 1306-1310 (1991)
MEDLINE 91142161
COMMENT Centromere cen2 tRNA cluster, B-repeat region 4 DNA.
FEATURES
    source
        1..1489
            location/Qualifiers
                /organism="Schizosaccharomyces pombe"
                /db_xref="taxon:4896"
                /gene="Val-tRNA"
                /note="codon recognized: GTT"
                /product="tRNA-Val"
                /anticodon="(pos:267..269,aa:Val)"
                /join(233..270,281..315)
                /gene="Val-tRNA"
                /note="codon recognized: ATC"
                /product="tRNA-Val"
                /anticodon="(pos:35..37,aa:Val)"
                /number=1
                /join(233..270,281..315)
                /gene="Val-tRNA"
                /note="does not fit consensus"
                /cons-splice="(5'site:no,3'site:no)"
                /number=2
                /gene="Val-tRNA"
                /note="codon recognized: GCT"
                /product="tRNA-Ala"
                /anticodon="(pos:1010..1012,aa:Ala)"
                /number=1049
                /gene="Ala-tRNA"
                /note="complement(1123..1196)"
                /gene="Ile-tRNA"
                /note="codon recognized: CTT"
                /product="tRNA-Ile"
                /anticodon="(pos:1157..1159,aa:Ile)"
                /complement(1123..1196)
                /gene="Ile-tRNA"
BASE COUNT 541 a 176 c 232 g 540 t
ORIGIN
Query Match 78.9%; Score 15; DB 27; Length 1489;
Best Local Similarity 94.1%; Pred. No. 1.32e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 366 CTGCTTAATTAAGAT 382
|||||

```

OY 1 CTGCTTAATTAAGAT 17

```

RESULT 5 YSPGTGIVAB 1489 bp DNA PLN 28-APR-1995
LOCUS Schizosaccharomyces pombe Ile-1, Ala-1 and Val-1 tRNA genes.
DEFINITION M57698
ACCESSION 9173496
MID 9173496
VERSION M57698.1 GI:173496
KEYWORDS transfer RNA; transfer RNA-Ala; transfer RNA-Ile-I; transfer RNA-Val-I.
SOURCE Schizosaccharomyces pombe DNA.
ORGANISM Schizosaccharomyces pombe
Eukaryotae; mitochondria; eukaryotes; Fungi; Ascomycota;
Archaeoscomycetes; Schizosaccharomycetales;
Schizosaccharomycetaceae; Schizosaccharomyces.
REFERENCE 1 (bases 1 to 1489)
AUTHORS Kuhn,R.M., Clarke,L. and Carbon,J.
TITLE Clustered tRNA genes in Schizosaccharomyces pombe centromeric DNA
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88 (4), 1306-1310 (1991)
MEDLINE 91142161
COMMENT Centromere cen2 tRNA cluster, B-repeat region 3 DNA.
FEATURES
    source
        1..1489
            location/Qualifiers
                /organism="Schizosaccharomyces pombe"
                /db_xref="taxon:4896"
                /gene="Ile-tRNA"
                /note="codon recognized: ATT"
                /product="tRNA-Ile"
                /anticodon="(pos:328..330,aa:Ile)"
                /number=367
                /gene="Ile-tRNA"
                /note="complement(441..514)"
                /product="tRNA-Ala"
                /note="codon recognized: CAT"
                /product="tRNA-Ala"
                /anticodon="(pos:475..477,aa:Ala)"
                /complement(441..514)
                /gene="Ala-tRNA"
                /note="codon recognized: CTT"
                /product="tRNA-Val"
                /anticodon="(pos:1175..1210,1220..1257)"
                /complement(1175..1210)
                /gene="Val-tRNA"
                /number=2
                /complement(1175..1257)
                /gene="Val-tRNA"
                /note="codon recognized: CGA"
                /product="tRNA-Val"
                /anticodon="(pos:1209..1211,aa:Val)"
                /complement(1175..1210,1220..1257))
                /gene="Val-tRNA"
                /number=1219
                /gene="Val-tRNA"
                /note="does not fit consensus"
                /cons-splice="(5'site:no,3'site:no)"
                /complement(1220..1257)
                /gene="Val-tRNA"
BASE COUNT 540 a 232 c 176 g 541 t
ORIGIN
Query Match 78.9%; Score 15; DB 27; Length 1489;
Best Local Similarity 94.1%; Pred. No. 1.32e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 1108 ATGCTTAATTAAGAG 1124
|||||
CP 17 ATGCTTAATTAAGAG 1

```



LOCUS	6	VSPTGIVAC	2303 bp	DNA	PLN	28-APR-1995
DEFINITION		S.pombe	Lys-1, Ile-1, Ala-1, Val-1 and Glu-2	trNA genes.		
ACCESSION		M57700				
NID		9173498				
VERSION		M57700.1	GI:173498			
KEYWORDS		transfer RNA; transfer RNA-Ala; transfer RNA-Glu-II; transfer RNA-Ile-1; transfer RNA-Lys-1; transfer RNA-Val-1.				
SOURCE		Schizosaccharomyces pombe DNA.				
ORGANISM		Schizosaccharomyces pombe				
REFERENCE		Eukaryote; mitochondrial eukaryotes; Fungi; Ascomycota; Archaeascomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomyces.				
AUTHORS		1 (bases 1 to 2303)				
TITLE		Kuhn, R.M., Clarke, L. and Catbon, J.				
JOURNAL		Clustered trNA genes in Schizosaccharomyces pombe centromeric DNA				
MEDLINE		sequence repeats				
COMMENT		Proc. Natl. Acad. Sci. U.S.A. 88 (4), 1306-1310 (1991)				
FEATURES		91142161				
Source		Centromere cen2 trNA cluster, B-repeat region 5 DNA.				
trNA		1. .2303				
		/organism="Schizosaccharomyces pombe"				
		/db_xref="taxon:4896"				
		join(303. .340,349. .384)				
		/gene="Lys-trNA"				
		/note="codon recognized: TAT"				
		/product="trNA-Lys"				
		/anticodon="(pos:337. .339,aa:Lys)				
		join(303. .340,349. .384)				
		/gene="Lys-trNA"				
		341. .348				
		/gene="Lys-trNA"				
		/note="does not fit consensus"				
		/cons_splice="(5' site:no,3' site:no)				
		349. .384				
		/gene="Lys-trNA"				
		/number=2				
		682. .755				
		/gene="Ile-trNA"				
		/note="codon recognized: ATT"				
		/product="trNA-Ile"				
		/anticodon="(pos:716. .718,aa:Ile)				
		682. .755				
		/gene="Ile-trNA"				
		complement(829. .902)				
		/gene="Ala-trNA"				
		/note="codon recognized: CAT"				
		/product="trNA-Ala"				
		/anticodon="(pos:863. .865,aa:Ala)				
		complement(829. .902)				
		/gene="Ala-trNA"				
		complement(1562. .1645)				
		/note="codon recognized: GCG; precursor"				
		/product="trNA-Val"				
		/anticodon="(pos:1596. .1598,aa:Val)				
		complement(1562. .1598)				
		/gene="Val-trNA"				
		/number=2				
		complement(join(1562. .1598,1608. .1645))				
		/gene="Val-trNA"				
		/note="codon recognized: TAT"				
		/product="trNA-Val"				
trNA						
exon						
intron						

```

gene      /anticodon=(pos:35..37,aa:Val)
          complement(join:1562..1598,1608..1645))
          /gene="Val-tRNA"
          1599..1607
          /gene="Val-tRNA"
          /note="does not fit consensus"
          /cons.splice=(5/site:no,3/site:no)
          complement(1608..1645)
          /gene="Val-tRNA"
          /number=1
          complement(2086..2157)
          /gene="Glu-tRNA"
          /note="codon recognized: GGT"
          /product="tRNA-Glu"
          /anticodon=(pos:2119..2121,aa:Glu)
          complement(2086..2157)
          /gene="Glu-tRNA"

BASE COUNT      819 a      375 c      321 g      787 t      1 others

ORIGIN

Query Match      78.9%; Score 15; DB 27; Length 2303;
Best Local Similarity 94.1%; Pred. No. 1,326+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1496 ATCCGTTATTAATGAAG 1512
      ||||| ||||| |||||
Cp 17 ATCGTTCATTATGAAG 1

RESULT LOCUS 7
DEFINITION EHU04863 2639 bp mRNA INV 29-NOV-1995
          Entamoeba histolytica HMI:IMSS alcohol dehydrogenase 2 (EhADH2)
          mRNA, complete cds.
ACCESSION U04863
          9488429
NID U04863.1 GI:488429
KEYWORDS Entamoeba histolytica.
SOURCE Entamoeba histolytica.
          Eukaryote; mitochondrial eukaryotes; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 2639)
          Yang,W., Li,E., Kairong,T. and Stanley,S.L. Jr.
          Entamoeba histolytica has an alcohol dehydrogenase homologous to
          the multifunctional adhe gene product of Escherichia coli
          Mol. Biochem. Parasitol. 64 (2), 253-260 (1994)
          95021502
          2 (bases 1 to 2639)
          Stanley,S.L.
          Direct Submission
          Submitted (06-JAN-1994) Samuel L. Stanley, Washington University
          School of Medicine, Medicine/Infectious Diseases, 660 S. Euclid
          Avenue, St. Louis, MO 63110, USA

FEATURES
Source Location/Qualifiers
          1..2639
          /organism="Entamoeba histolytica"
          /strain="HMI:IMSS"
          /db_xref="taxon:5759"
          3..2615
          /standard_name="EhADH2"
          /note="The derived amino acid sequence of EhADH2 is
          homologous (48% identity) to the trifunctional enzyme
          product of the E. coli adhe gene described by Goodlove et
          al., Gene 85:209-214, 1989 (Swiss Prot Accession Number
          P11547)"
          /codon_start=1
          /evidence=experimental
          /product="alcohol dehydrogenase 2"
          /protein_id="AA81906.1"
          /db_xref="PID:9488430"
          /db_xref="GI:488430"
          /translation="MSTQQTMTVDENHNLQVLRKAQVALKEIKETFOEKIDYIVKRA
          SVALADHCAAAAAVEETGTEGTEEDATNINFCETVHEMRRAKTVGIINDPLPG
          ITTAEAVGAAVCGTPTNPTNPAIFKSLISIKIRNDIVSFHPSALKCSIMAKIVR

```



DB 831 GCATATTTCATTATGAAG 849  
 CP 19 GCATCGTTCATTATGAAG 1  
 RESULT 8  
 LOCUS EHNADDAD 2652 bp mRNA INV 11-MAY-1995  
 DEFINITION E.histolytica mRNA for NAD+-dependent alcohol dehydrogenase.  
 ACCESSION X77132  
 NID 5443984  
 VERSION X77132.1 GI:443984  
 KEYWORDS alcohol dehydrogenase; NAD+-dependent alcohol dehydrogenase.  
 SOURCE Entamoeba histolytica.  
 ORGANISM Eukaryota; Entamoebidae; Entamoeba.  
 REFERENCE 1 (bases 1 to 2652)  
 Bruchhaus, I. and Tennich, F.  
 Purification and molecular characterization of the NAD(+) dependent  
 acetaldehyde/alcohol dehydrogenase from Entamoeba histolytica  
 Biochem. J. 303 (Pt 3), 743-748 (1994)  
 JOURNAL MEDLINE 95071285  
 REFERENCE 2 (bases 1 to 2652)  
 Bruchhaus, I.  
 Direct Submission  
 Submitted (12-JAN-1994) I. Bruchhaus, Bernhard Nocht Institute for  
 TROPICAL Medicine, Bernhard-Nocht-Str. 74, 20359 Hamburg, FRG  
 JOURNAL MEDLINE  
 FEATURES  
 SOURCE  
 1. 2652  
 /organism="Entamoeba histolytica"  
 /note="pathogen"  
 /strain="HM-1:IMSS"  
 /db\_xref="taxon:5759"  
 /dev\_stage="trophozoite"  
 /clone\_lib="lambda ZAP"  
 1..2619  
 /gene="ADH2"  
 1..2619  
 /gene="ADH2"  
 /EC\_number="1.1.1.1"  
 /note="alcohol dehydrogenase"  
 /product="NAD+-dependent alcohol dehydrogenase"  
 /protein\_id="CAA54388.1"  
 /db\_xref="PID:g443985"  
 /db\_xref="GI:443985"  
 /db\_xref="SWISS-PROT:O24803"  
 /translation="MMSSTQGTMTVDHINQVLAQAVALKETIKPEYTEKIDYIVK  
 KASVAALDQCALAAVAEETGRGIFEDKATKINIFACEVTHEHRAKTVGIIINVDL  
 YGITEIAEPGVGVGVPTNPSTAFKSLISIKTRNPIVFSFHSALCKSIIMAKI  
 VRDAIAGAAPENCIOIWEFGIEASRKLNNHGVATILATGNAAMKAYSSGKPAI  
 VGAGANVPTIEKTCNKIOAANDVMSKSFSDNGICASEDAALIDKEIYQVVEKMT  
 LGAVFNEEKAKLEKFMFGVNAVSADVNNAIRNPKCPGMSQPMFAEONGIKYVPEDCN  
 ICAVCKEVPNEPIREKLSPLAIIKAENTODGIDKAAVNEFNGRSHSAIHSND

DB 835 GCATATTTCATTATGAAG 853  
 CP 19 GCATCGTTCATTATGAAG 1  
 RESULT 9  
 LOCUS SPCNZB 3044 bp DNA PLN 06-JUL-1993  
 DEFINITION S.pombe cent central region centromere DNA (pYC113).  
 ACCESSION X66740  
 NID 9313341  
 VERSION X66740.1 GI:313341  
 KEYWORDS centromere; centromere 2; transfer RNA; transfer RNA-Ala; transfer  
 RNA-Ile; transfer RNA-Val.  
 SOURCE fission yeast.  
 ORGANISM Schizosaccharomyces pombe  
 Eukaryota; Fungi; Ascomycota; Archiascomycetes;  
 Schizosaccharomycetales; Schizosaccharomycetaceae;  
 Schizosaccharomyces.  
 REFERENCE 1 (bases 1 to 3044)  
 Takahashi, K.  
 Direct Submission  
 Submitted (16-JUN-1992) K. Takahashi, Dept of Biophysics, Faculty  
 of Science, Kyoto University, Sakyo-ku Kyoto 606, JAPAN  
 JOURNAL MEDLINE 92385892  
 REFERENCE 2 (bases 1 to 3044)  
 Takahashi, K., Murakami, S., Chikashige, Y., Funabiki, H., Niwa, O. and  
 Yanagida, M.  
 A low copy number central sequence with strict symmetry and unusual  
 chromatin structure in fission yeast centromere  
 Mol. Biol. Cell 3 (7), 819-835 (1992)  
 JOURNAL MEDLINE  
 FEATURES  
 SOURCE  
 1. 3044  
 /organism="Schizosaccharomyces pombe"  
 /strain="972"  
 /db\_xref="taxon:4896"  
 /chromosome="II"  
 misc\_feature 1..80  
 /note="homology with dhia"  
 tRNA 1856..1929  
 /product="transfer RNA-Ile"  
 tRNA 2003..2076  
 /product="transfer RNA-Ala"  
 tRNA 2676..2750  
 /product="transfer RNA-Val"  
 BASE COUNT 1077 a 486 c 414 g 1058 t 9 others  
 ORIGIN  
 Query Match 78.9%; Score 15; DB 21; Length 3044;  
 Best Local Similarity 94.1%; Pred. No. 1.32e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 DB 2663 ATCGTATTAATGAAG 2679  
 CP 17 ATCGTTCATTATGAAG 1  
 RESULT 10



LOCUS ANANFJ 3866 bp DNA BCT 24-NOV-1993  
DEFINITION Anabaena sp. (PCC 7120) nitrogen fixation protein (nifH) gene,  
complete cds.  
ACCESSION L14925  
NID 9289138  
VERSION L14925.1 GI:289138  
KEYWORDS  
SOURCE Anabaena sp. (strain PCC 7120) DNA.  
ORGANISM Anabaena sp.  
REFERENCE 1 (bases 1 to 3866)  
AUTHORS Bauer, C.C., Scapino, L. and Haselkorn, R.  
TITLE Growth of the cyanobacterium Anabaena on molecular nitrogen: NifH  
is regulated when iron is limited  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (19), 8812-8816 (1993)  
MEDLINE 94022264  
REFERENCE 2 (bases 1 to 3866)  
AUTHORS Bauer, C.C.  
TITLE Direct Submission  
JOURNAL Submitted (28-APR-1993) Christopher C. Bauer, Molecular Genetics  
and Cell Biology, University of Chicago, Chicago, IL 60637 USA  
LOCATION/Qualifiers  
1. 3866  
/organism="Anabaena sp."  
/strain="PCC 7120"  
/db\_xref="taxon:1167"  
/gene="nifH"  
88..3681  
/gene="nifH"  
88..3681  
/gene="nifH"  
/codon\_start=1  
/transl\_table=11  
/product="nitrogen fixation protein"  
/protein\_id="AAC36818.1"  
/db\_xref="PID:q289139"  
/db\_xref="GI:289139"  
/translation="MSQTRATIDGNEANAVAKYKLENEVATITPTSSAMGEMADAM  
AGRPMLGTVPSSVVOSEGAAGVHALDTGISTFTASQGLLMIPLNLYKIG  
ELTSMVHVAAKSLATHALSIFGSDHVAAGATGFMALCSASVESHDFALIAHAT  
LDTKSVFLHFDGFTSHVQVELLADDDVSLINEDKIFAHRAALTPDPLRGAT  
ANPDVFEQARGANPYNACPAIVQIGDKFGERTGYOIEYHGASDRLIIT  
MSGCTVHETVDYLNARGEKGVKRLFRPMDVPRVGOALPHSOALVADRIKPE  
GSAGEPLVDVYTAIHGKVNKNNSVPSQSPVRIIGRIGLSKETPTMAKAVE  
DNLAQTPKRNHFTTIGINDVTRTSLTSDSTEDPNVRAVFGSGTGVANKNLS  
KILTEGIDNVAQGFVYDSKSSMTVSLRFGSPISRTYLIDANFIGHNAFL  
ELIEVNLRLMATILLNSPYNAITWENPLKVRQIILDKOLKLYINNOVARDSGM  
GRIINTMOVCFEPALAGVLPVOALAKIKOALEKTYGKGVEMVNIQAADVOTLENT  
HEVKIPIEKGRKIDPEALLSNQSPSTSPAFKRVLYCKINWOGDDIPVLTLPDGG  
TFPTGAKEKKNVAOEIPVWDITICVQCSKVCVCPHAIKAYIQPSELNAPTE  
KSVDAKDRDFANOKETIOVAPEDCTGCALCVCPKANKSESLKAINANOLDPLREO  
KSDNDPFLNLPNDRNLKLNQIRQOQOEPLFERSGACAGGETPYKLLTOLFGD  
RSVINATGCSISGNLPTTPMTKNNDRGPANSLFEDNAEEFGYKRLSDQAE  
FAELLQOSTEVDNLVDSILKAPQKTEADIMEORRIELIKOOLDIPTPDNPKS  
KRONLSADYLVYKRSVAILGGDMAVDDFGSIDRYIASGRNVILNVDIYVSTG  
GSSKATPFAAFAKFAKPAOKKMDGMANVYVAVSALGAKDOOTLKALEA  
EAFDGSIIIAVSHCIANGINMTGNNOOKALVESGRMLYRNPLDLOGKPNPLD  
MRSPQSVOSMYQENRFKMLTKRSKEVAKOLLEQAEVDARWMTQYIASR"

repeat\_unit  
/gene="nifH"  
1164..1195  
/note="heptamers (CCCCAGT) x 5"  
repeat\_unit  
/note="heptamers (CCCCAGT) x 5"  
/P1L\_type=tandem  
3677..3710  
/note="heptamers ((A/G)GGTAAAT) x 5"  
/rpt\_type=tandem  
BASE COUNT 1099 a 835 c 907 g 1025 t  
ORIGIN  
Query Match 78.9%; Score 15; DB 18; Length 3866;  
Best Local Similarity 94.1%; Pred. No. 1.32e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 663 ATCGTCATTAATGAAG 679  
||||| |||||||||

Cp 17 ATCGTCATTAATGAAG 1  
RESULT 11  
LOCUS TAUS1330 3878 bp DNA PLN 07-FEB-1997  
DEFINITION Triticum aestivum leaf rust resistance kinase Lr10 (LrK10) gene,  
complete cds.  
ACCESSION U51330  
NID 91680685  
VERSION U51330.1 GI:1680685  
KEYWORDS  
SOURCE wheat.  
ORGANISM Triticum aestivum  
REFERENCE 1  
AUTHORS Feuillet, C., Schachemayr, G. and Keller, B.  
TITLE Molecular cloning of a new receptor-like kinase gene encoded at the  
Lr10 disease resistance locus of wheat  
JOURNAL Plant J. 11 (1), 45-52 (1997)  
MEDLINE 97177795  
REFERENCE 2 (bases 1 to 3878)  
AUTHORS Feuillet, C.  
TITLE Direct Submission  
JOURNAL Submitted (14-MAR-1996) Catherine Feuillet, Plant Breeding, Swiss  
Federal Research Station for Agronomy, Reckenholzstr. 191, Zurich  
8046, Switzerland  
LOCATION/Qualifiers  
1. 3878  
/organism="Triticum aestivum"  
/db\_xref="taxon:4565"  
/chromosome="1"  
/tissue\_type="leaf"  
619..3487  
/gene="LrK10"  
join(619..1432,2270..2293,2415..3487)  
/gene="LrK10"  
/transl\_table=1  
/product="rust resistance kinase Lr10"  
/protein\_id="AAC49629.1"  
/db\_xref="PID:q1680686"  
/db\_xref="GI:1680686"  
/translation="MSKLLVLAALLPLINHGIVLATAMDQDFEYKCPSPKSGQGP  
MIRYPLGESSNTSSSSCGAGRSIMWLACSGDPTIVHPVLPYSVSAIDYRSM  
KITPLVDCLIVLOKLLISRSSSPQVAVINDEKSPENFESSATIVGSRFTT  
AAAHDDIAGVPVSCISNTHFXYVNSDEMSILPDKVYVSPRGISLPHMKDO  
MFYNTETAKKIPSAETAVSMBDEDCRECELSGRCAFPSSORDREFCMPDHSHIK  
VIAATSSVAAPVALLTYATVLYLKLTRYNAELIMKVMELKTYGTGTPRTYFSEV  
KMAARFEKVGOGGFGVYKGLNGVAVKMDNSTGESEFINVATIGLTHAA  
NIVRLIGSEEMRALIYEFMPNLSLEYFSDSNIFONLVEKLLDIALGARG  
MEYLHOGNORLTHEDIRPHNLLDYNENPKISDGLAKLAROOSITVTLARBTMG  
YIAPLPLSRNMGVSKADYVSFGMLVEMSGRNSPRISODVILPENTYKVI  
NQEELALLETTOEKDKVRQOLAMALCLIONPKNRPSMTKVMNLGRLQSLQMP  
KPFVSESELMNS"

BASE COUNT 1142 a 833 c 831 g 1072 t  
ORIGIN  
Query Match 78.9%; Score 15; DB 28; Length 3878;  
Best Local Similarity 94.1%; Pred. No. 1.32e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 2707 ATCGTCATTAATGAAG 2723  
||||| ||||||| |||||||  
Cp 17 ATCGTCATTAATGAAG 1  
RESULT 12  
LOCUS YSCRAD2G 3919 bp DNA PLN 17-MAY-1995  
DEFINITION S.cerevisiae RAD2 gene, complete cds.  
ACCESSION M10275  
NID 9172328  
VERSION M10275.1 GI:172328



[illegible]

/db\_xref="PID:q172329"  
 /db\_xref="GI:172329"  
 /translation="MGVHSFWDIAGPTAPRVLTSELDKRAVADASIWIYQTLKAVD  
 QEGNAVKNSHITGFFRRICLLYFGIRPVVFGVGVYLREPIRORKEERQKRESA  
 KSTARKLALATQNGSNDVKNSTSPSSGSSVQLFKPDMDMLDPIPGFKYKDEARVN  
 SNKFEKELMNSINDGLIEDIDTINPASADEFELPKATQYLLISLRATSRIMGYS  
 KEQLETPNSMDSERROIDMVKRRNFQKTLINTGFCGGGSKLINEPIINTSGK  
 SKETKTLKTNGWITGLIGANDGSDQAIAIYDDDAALAKKQDLSNAGEDVLRMDL  
 EDNSLKTIVREHSSNATPAQKRSRNSRDECDCEWEVEELPKRVKVEFEFSLKA  
 ARIPMGOSLNNAGSSKSFILDRHDOASPSKTPTPMRSISVEDEDDYKQIIEILM  
 MEAPVLSKMEKKEPPADCKSKIAKVPYTKGTEARPIYQYGLGDDPSKQPYHYTNLN  
 SKSSVLSKSKTKTYLSEFRPPOCEDKALITLGEQNLINTSHKIPQFDNNNSLIF  
 OKNTESVSOEATKEKPIPIEMBSMESTASQOLYNNTNTNYEDKNNVEDSCAE  
 ITNGSSYELLTGHNATELIERESKSSSNDENKDDLEVLSELEFDPVTKSQISKE  
 AEEDDSKRSVSEINKENRRPLIFDYFSEDEEDINIVEMKIEQEEFDFKNTLSTSAE  
 RNVAENAFVEDELFEQOMKDRSDSEVTMMKIEVOLLSRFGIPTATPMEAQCA  
 ELIOLNIVDGIIDELSDVFLFEGGKRIYKNNEHENYVEFADASTILKGLDKRNMIE  
 LAOLGSDTYNGLKGMKPVSSIVIAFGLKNNKPMYNNNGQDKRKQEIENKFEIDL  
 RKIKVNNETIILDDDFPSVWYDAIMRPEVHDITPFWAGVPDMLMSEFKTDLGWHK  
 EKSDEILPIPIRVNKKRRKKQKRIINEFPREVISGDKRLNTSRISTATGKLKKRK  
 M"  
 3445..3446  
 /gene="RAD2"  
 /citation="[1]  
 BASE COUNT 1355 a 653 c 908 g 1003 t  
 ORIGIN 223 bp upstream of Ndel site [1]: chromosome 7 [1].  
 Query Match 78.9%; Score 15; DB 27; Length 3919;  
 Best Local Similarity 89.5%; Pred. No. 1,332+0; Indels 0; Gaps 0;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 3450 GCCTGCTTCATTCATGAAG 3468  
 |||||||  
 Cp 19 GCATGCTTCATTAAGAAG 1

RESULT 13  
 LOCUS SCYGR258C 5175 bp DNA PLN 11-AUG-1997  
 DEFINITION S.cerevisiae chromosome VII reading frame ORF YGR258C.  
 ACCESSION Z73043 Y13135  
 NID G1323469  
 VERSION Z73043.1 GI:1323469  
 KEYWORDS  
 SOURCE  
 ORGANISM baker's yeast.  
 Saccharomyces cerevisiae  
 Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 Saccharomycetaceae; Saccharomycetes.  
 1 (bases 1 to 603)  
 Agostoni Carbone,M.L., Panzeri,L., Melchiorretto,P., Carignano,I.G.,  
 Feroli,F., Frontali,L., Mazzoni,C., Rinaldi,T. and Ruzzi,M.  
 Unpublished  
 2 (bases 291 to 5175)  
 Panzeri,L., Agostoni Carbone,M.L., Melchiorretto,P., Plevani,P.,  
 Martegani,E., Vannoni,M., Carignano,I.G., Clemente,M.L., Frontali,L.,  
 Fabbiani,L., Marconi,A., Ruzzi,M. and Salicrú,M.  
 Unpublished  
 3 (bases 1 to 5175)  
 MIPS.  
 Direct Submission  
 Submitted (14-MAY-1996) Data collected by MIPS on behalf of the  
 European yeast chromosome VII sequencing project. MIPS at the  
 Max-Planck-Institut fuer Biochemie, Am Klopfersplitz 18a D-82152  
 Martinsried, FRG; E-mail: Mewes@mips.emblnet.org  
 Location/Qualifiers  
 1..5175  
 /organism="Saccharomyces cerevisiae"  
 /db\_xref="taxon:4932"  
 /chromosome="VII"  
 complement(366..3461)  
 /gene="RAD2"  
 /db\_xref="SGD:L0001556"  
 complement(366..3461)  
 /gene="RAD2"

CDS  
 gene  
 FEATURES  
 source



/note="ORF YGR258c"  
/codon\_start=1  
/db\_xref="SGD:L0001556"  
/protein\_id="CA97287.1"  
/db\_xref="PID:e243580"  
/db\_xref="PID:g1323470"  
/db\_xref="GI:1323470"  
/db\_xref="SWISS-PROT:P07276"  
/translation="MGVHSEWDIAGPTAPVRVLEEDKRMADVASIMVYIOLKAVRD  
OEGNAVNKSHITGFERRICKLYFGIRPYVEEDGVVPLKREIRORRERROGRRESA  
KSTARLKALALONGSNDVKNSTPSSGSSVOIEFPODEMPLDPIRPFYDDEDAVN  
SKTPEKLNISNGDLEDIDITINPAHEPELEKAKOYLILSLRLKRLMRYTS  
KDELTFPNNSMDFSRQIDVAKRNFFQKLNITGFDGGASKLNEEVINISQK  
SKEYLTKNTNGWILGLANDSGDAQKALVIDKDGALVKQLDSNAEDGDLRWMDL  
ENDSLKIVHESNNATAPQKRSRSEDCDECEMEVELKPNVFEVDSIKMA  
ARLPYGOSLNNAGSKFLDRHDOASPKTPTMRISISVEDDEDYIKOIEEEM  
MEAVOLSKMEKREAPDADSKIAKPVTSKGTETAPVVOYGLGAOPDSKOPHTNLTN  
SKSEVYIKRTSKTVLSEFPSPQOEDKGAILEGQNTLNPIFSKIPROPPNNNSILF  
QKNTBNSVQDEATKEKSPPEMPSMFSSTASQOLINPYTNTNVEDKVRNDESGAE  
TNNKGSYELLTGLNATEILERESEKSSNDKDDLEVLSEELFEDVPTKSQISKE  
AADNSRKVESINKERKPLIDYDFSEDEENIYVNMKEOEEDPTFKNTLSISAE  
RVNAENAFVEDELEEQOMKRDSDDEVIMDKIEVQLSRGIPYITAPMAEAOCA  
ELLQNLVVGITDSDVLFEGGTIKYKMPFEKNVVERVDAESILKLGDRKNMIE  
LAOLGSDYTNGLKMGKPYSTIEVTAEPGNLKNFKRMVYNNQOPDKRKQETENKFEEDL  
RKLYNNEIILDDPPSPVAYDIYMRPEVDHDTTPVWCVLPDLMLRSTKLTOLGMPH  
EKSDLEIPLIRDVNKKRKKQKRIINEFPREYISGDKRLNTSKRISTATGKLKRRK  
M"

BASE COUNT 1332 a 1186 c 870 g 1787 t

ORIGIN

Query Match 78.9%; Score 15; DB 27; Length 5175;  
Best Local Similarity 89.3%; Pred. No. 1.32e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 567 CTCATGATGAAGACGAGC 585  
||||| ||||||| ||  
QY 1 CTCATTATGAACGATGC 19

## RESULT 14

LOCUS AF005098 5565 bp DNA BCT 23-MAR-1998  
DEFINITION Lactococcus lactis RNaseH II (rnhb) gene, partial cds, positive  
regulator Gadr (gadr), Gadc (gadc) and glutamate decarboxylase  
(gadb) genes, complete cds.

ACCESSION AF005098  
NID G2352483  
VERSION AF005098.1 GI:2352483  
KEYWORDS  
SOURCE  
ORGANISM

Lactococcus lactis.  
Lactococcus lactis  
Eubacteria; Firmicutes; Low G+C gram-positive bacteria;  
Streptococcaceae; Lactococcus.

REFERENCE 1 (bases 1 to 5565)  
AUTHORS Sanders, J.W., Leenhouts, K., Burghoorn, J., Brands, J.R., Venema, G.  
and Kok, J.

TITLE A chloride-inducible acid resistance mechanism in Lactococcus  
lactis and its regulation  
JOURNAL Mol. Microbiol. 27 (2), 299-310 (1998)

REFERENCE 2 (bases 1878 to 2423)  
AUTHORS Sanders, J.W., Venema, G., Kok, J. and Leenhouts, K.  
TITLE Identification of a sodium chloride-regulated promoter in  
Lactococcus lactis by single-copy chromosomal fusion with a  
reporter gene

JOURNAL Mol. Gen. Genet. 257 (6), 681-685 (1998)  
MEDLINE 98265928

REFERENCE 3 (bases 1 to 5565)  
AUTHORS Sanders, J.W., Venema, G., Kok, J. and Leenhouts, K.  
TITLE Direct Submission

JOURNAL Submitted (20-MAY-1997) Genetics, Groningen Biomolecular Sciences  
and Biotechnology Institute, University of Groningen, Kercklaan 30,  
Haren 9715 NN, The Netherlands  
FEATURES Location/Qualifiers

source 1..5565  
/organism="Lactococcus lactis"  
/strain="M31363"  
/db\_xref="taxon:1358"  
/map="between 1dh and leu-11v"  
/clone="NS3"  
<1..555  
/gene="rnhb"  
/gene="rnhb"  
<1..503  
/gene="rnhb"  
/CDS\_number="3.1.26.4"  
/function="specifically degrades the ribonucleotide moiety  
on RNA-DNA hybrids"  
/note="Rnhb; similar to Escherichia coli RNaseH II"  
/codon\_start=3  
/transl\_table=11  
/product="RNaseH II"  
/protein\_id="AAC46185.1"  
/db\_xref="PID:g2352484"  
/db\_xref="GI:2352484"  
/translation="AVILPKNCKIRGLNDSKKVPKSKHAILSEIQERKALAGVIVD  
AKSIDENVYEATKIAMIOAVSKLSKPEHLIDAMVLDLPIAOTRKIHGDRSSIA  
ASAIYAKYTRDMMKDFLPEPPEYFENNAGYGTAKHIALTKYGTIRHRSYEPK  
SNVNRK"  
526..555  
/gene="rnhb"  
941..946  
/gene="gadr"  
941..1967  
/gene="gadr"  
965..969  
/gene="gadr"  
979..1966  
/gene="gadr"  
1084..1089  
/gene="gadr"  
1095..1925  
/gene="gadr"  
/function="activates chloride dependent transcription of  
gadb"  
/codon\_start=1  
/transl\_table=11  
/product="positive regulator Gadr"  
/protein\_id="AAC46186.1"  
/db\_xref="PID:g2352485"  
/db\_xref="GI:2352485"  
/translation="MKKYGDCFKRLRNQKNIGLSYFSLGIDRSNISRFEHGKCMS  
FERIDMLEKQVPLSEYELIVNFMPPNPOEFFILELEKAEFSQNRDKIKELYSVKE  
TGNHLITVTKTKLIGNISQTEVKEIEALVCNIEKNGYFELTFYVSQYLVNNOLEL  
LNFPRKRCNVCRLVKYRRRLQIAYKSAIYAAGERKKAENILEMTRKYRTVGVDL  
YSEVLRHLARAILIFNEFAEIGEKEKNVLALEILPEFGKKIKKEYQNMKEVLRASI  
"

stem\_loop 1926..1967  
/gene="gadr"  
/note="functions as terminator; may function in chloride  
dependent transcription"  
1987..3580  
/gene="gadc"  
1987..1992  
/gene="gadc"  
2001..5037  
/gene="gadc"  
2058..2065  
/gene="gadc"  
2069..3580  
/gene="gadc"  
/note="Putative glutamate - gamma-aminobutyrate  
antiporter; Functions in glutamate dependent acid stress  
resistance. Expression induced by chloride and at low pH.  
Homologue of Shigella flexneri Gadc"  
/codon\_start=1  
/transl\_table=11  
/product="Gadc"  
/protein\_id="AAC46187.1"







\*\*\*\*\*  
MIRCH  
\*\*\*\*\*

(TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msrch\_ntp n.a. - n.a. Smith-Waterman search, using a protein database  
which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Nov 24 02:28:42 1999; Maspar time 0.37 Seconds

Tabular output not generated. 118,965 Million cell updates/sec

Title: >US-09-103-287-5  
Description: (1-19) from US09103287.seq  
Perfect Score: 95  
N.A. Sequence: 1 CTTCATTATGACGATGC 19  
Comp: GAAGTATTAATCTCTAGC

Scoring table: TABLE dkttranslated2  
Gap 40

Nmatch STD : Dbase 0; Query 0

Searched: 1479 segs, 1162203 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: HIV-AAB  
1:ALL

Statistics: Mean 37.456; Variance 38.083; scale 0.984

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	56.9	240	1	NEFS2D205	HIV2D205, NEF PROTEIN	5.63e+00
2	56.9	486	1	ENVELOPES4	SIVAGMTB14, ENV POLYP	5.63e+00
3	55.9	492	1	GAGSUG280	HIVG280, GAG POLYPR	8.11e+00
4	55.9	557	1	ENV_U08804	HIVU08804, ENV POLYPR	8.11e+00
5	55.9	768	1	ENVSAQM155	SIVAGM155, ENV POLYPR	8.11e+00
6	52.7	424	1	LOCUSSHMT	HUMAN IMMUNODEFICIENC	2.28e+01
7	52.7	860	1	ENV2BEN	HIV2BEN, ENV POLYPR	2.28e+01
8	51.7	27	1	ENVFLO5R4	HIVFLO5R4, ENV	3.15e+01
9	51.7	180	1	NEFS2N1H2	HIV2N1H2, NEF PROTEIN	3.15e+01
10	51.7	214	1	NEFS2N1H2	SIVANDGB1, NEF	3.15e+01
11	51.7	225	1	NEFS2UC1	HIV2UC1, NEF PROTEIN	3.15e+01
12	51.7	247	1	NEFS2M251	SIVAM251, NEF PROTEIN	3.15e+01
13	51.7	255	1	NEFS2GHI	HIV2GHI, NEF PROTEIN	3.15e+01
14	51.7	257	1	NEFS2D194	HIV2D194, NEF PROTEIN	3.15e+01
15	51.7	257	1	NEFS2BEN	HIV2BEN, NEF PROTEIN	3.15e+01
16	51.7	260	1	NEFS2M251	SIVAMP11, NEF PROTEIN	3.15e+01
17	51.7	260	1	NEFS2CAM2	HIV2CAM2, NEF PROTEIN	3.15e+01
18	51.7	261	1	NEFS2MPBU	SIVSMPBU, NEF PROTEIN	3.15e+01
19	51.7	261	1	NEFS2M251	SIVSMPBU, NEF PROTEIN	3.15e+01

20	51.7	262	1	NEFS2M142	SIVM142, NEF PROTEIN	3.15e+01
21	51.7	263	1	NEFS2MNE	SIVMNE, NEF PROTEIN	3.15e+01
22	51.7	263	1	NEFS2M32H	SIVM32H, NEF PROTEIN	3.15e+01
23	51.7	263	1	NEFS2M1A11	SIVM1A11, NEF PROTEIN	3.15e+01
24	51.7	309	1	NEFS2M44	SIVSM44, NEF PROTEIN	3.15e+01
25	51.7	502	1	NEFS2M251	SIVSMPBU, NEF PROTEIN	3.15e+01
26	51.7	851	1	ENV2D194	HIV2D194, ENV	3.15e+01
27	51.7	856	1	ENVELOPES2	HIV2321, ENVELOPE POL	3.15e+01
28	51.7	882	1	ENVELOPES3	SIVSTM, ENVELOPE POLY	3.15e+01
29	51.7	885	1	ENVELOPES3	SIVSM4, ENVELOPE PO	3.15e+01
30	51.7	886	1	ENVELOPES3	SIVSM9, ENVELOPE PO	3.15e+01
31	51.7	886	1	ENVSM209	SIVSM209, ENV POLYPR	3.15e+01
32	51.7	889	1	ENVSM209	SIVSM209, ENV POLYPR	3.15e+01
33	51.7	1009	1	ENVSM209	SIVSM209, ENV POLYPR	3.15e+01
34	50.6	92	1	ENV_U08673	HIVU08673, ENV POLYPR	4.29e+01
35	50.6	92	1	ENV_U08672	HIVU08672, ENV POLYPR	4.29e+01
36	50.6	92	1	ENV_U08677	HIVU08677, ENV POLYPR	4.29e+01
37	50.6	96	1	VRPSBRU	HIVBRU, VPR PROTEIN	4.29e+01
38	50.6	96	1	VRPSCP2	SIVCP2, VPR PROTEIN	4.29e+01
39	50.6	96	1	VRPSBGS3C	HIVBGS3C, VPR PROTEI	4.29e+01
40	50.6	96	1	VRPSNHS2	HIVNHS2, VPR POLYPR	4.29e+01
41	50.6	96	1	VRPSJREL	HIVJREL, VPR PROTEIN	4.29e+01
42	50.6	185	1	ENVSAQM155	SIVAGMTB53, ENV POLYPR	4.29e+01
43	50.6	500	1	GAGSUTB31	HIVUTB31, GAG POLYPR	4.29e+01
44	50.6	500	1	GAGSUTB31	HIVUTB31, GAG POLYPR	4.29e+01
45	50.6	569	1	ENV_U08793	HIVU08793, ENV POLYPR	4.29e+01

## ALIGNMENTS

RESULT 1	PRELIMINARY; PRT; 240 AA
ID NEFS2D205	
AC X16109; X61240;	
DT 25-OCT-1991	
DE HIV2D205, NEF PROTEIN	
FT PEPT 1	240
CC -1-5'-Y 3'-X	NEF PROTEIN
CC TRANSLATED USING PHASE 1	
SEQ SEQUENCE 240 AA; 27406 MW;	
Query Match	Score 56; DB 1; Length 240;
Best Local Similarity 64.3%;	Pred. No. 5.63e+00;
Matches 9; Conservative	4; Mismatches 1; Indels 0; Gaps 0;
Db f 1 k e	
Dt 312 yttatthaaagc 325	
Qy 1 CTTCATTATGAC 14	
Qt F I N E	
RESULT 2	PRELIMINARY; PRT; 486 AA
ID ENVELOPES4GATMB14	
AC M80208	
DT 09-APR-1993	
DE SIVAGMTB14, ENV POLYPROTEIN	
FT PEPT <1	>486
CC -1-5'-N 3'-N	ENV POLYPROTEIN
CC TRANSLATED USING PHASE 2	
SEQ SEQUENCE 486 AA; 55458 MW;	
Query Match	Score 56; DB 1; Length 486;
Best Local Similarity 33.3%;	Pred. No. 5.63e+00;
Matches 6; Conservative	8; Mismatches 4; Indels 0; Gaps 0;
Db 1 n n e s	
Dt 1092 yttatthaaagc 1109	
Qy 1 CTTCATTATGACGATG 18	
Qt F I N E R	
RESULT 3	



RESULT 10



ID NEFSMNDGB1 PRELIMINARY; PRT; 214 AA  
AC M27470; X15781;  
DE 20-DEC-1989  
DE SIYVNDGB1, NEF  
FT PEPT 1 214 NEF nef  
CC -1-5':Y 3':Y  
CC TRANSLATED USING PHASE 1  
SQ SEQUENCE 214 AA; 24258 MW;

Query Match 53.7%; Score 51; DB 1; Length 214;  
Best Local Similarity 61.5%; Pred. No. 3.15e+01;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db f i k e  
Dt 330 ytyathargar 342  
OY 1 CTTCATTATGAA 13  
QT F I N E

RESULT 11  
ID NEFS2UC1 PRELIMINARY; PRT; 225 AA  
AC L07625  
DE 11-MAR-1992  
DE HIV2UC1, NEF PROTEIN  
FT PEPT 1 225  
CC -1-5':Y 3':Y  
CC TRANSLATED USING PHASE 1  
SQ SEQUENCE 225 AA; 25680 MW;

Query Match 53.7%; Score 51; DB 1; Length 225;  
Best Local Similarity 61.5%; Pred. No. 3.15e+01;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db f i k e  
Dt 309 ytyathargar 321  
OY 1 CTTCATTATGAA 13  
QT F I N E

RESULT 12  
ID NEFSM251 PRELIMINARY; PRT; 247 AA  
AC M19499; M15897; M16125; Y00283; X06879; Y00294; Y00295; X06391; X06393;  
DE 01-MAR-1989  
DE SIYVM251, NEF PROTEIN  
FT PEPT 1 247  
CC -1-5':Y 3':Y  
CC TRANSLATED USING PHASE 1  
SQ SEQUENCE 247 AA; 28414 MW;

Query Match 53.7%; Score 51; DB 1; Length 247;  
Best Local Similarity 61.5%; Pred. No. 3.15e+01;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db f i k e  
Dt 363 ytyathargar 375  
OY 1 CTTCATTATGAA 13  
QT F I N E

RESULT 13  
ID NEFS2GH1 PRELIMINARY; PRT; 255 AA  
AC M30895  
DE 19-DEC-1989  
DE HIV2GH1, NEF PROTEIN  
FT PEPT 1 255  
CC -1-5':Y 3':Y  
CC TRANSLATED USING PHASE 1  
SQ SEQUENCE 255 AA; 29248 MW;

Query Match 53.7%; Score 51; DB 1; Length 255;

Best Local Similarity 53.3%; Pred. No. 3.15e+01;  
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
Db f i k k  
Dt 363 ytyathargar 377  
OY 1 CTTCATTATGACG 15  
QT F I N E

RESULT 14  
ID NEFS2D194 PRELIMINARY; PRT; 257 AA  
AC J04542; X52223;  
DE 12-MAR-1991  
DE HIV2D194, NEF PROTEIN  
FT PEPT 1 257  
CC -1-5':Y 3':Y  
CC TRANSLATED USING PHASE 1  
SQ SEQUENCE 257 AA; 29448 MW;

Query Match 53.7%; Score 51; DB 1; Length 257;  
Best Local Similarity 61.5%; Pred. No. 3.15e+01;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db f i k e  
Dt 363 ytyathargar 375  
OY 1 CTTCATTATGAA 13  
QT F I N E

RESULT 15  
ID NEFS2BEN PRELIMINARY; PRT; 257 AA  
AC M30502  
DE 30-OCT-1989  
DE HIV2BEN, NEF PROTEIN  
FT PEPT 1 257  
CC -1-5':Y 3':Y  
CC TRANSLATED USING PHASE 1  
SQ SEQUENCE 257 AA; 29911 MW;

Query Match 53.7%; Score 51; DB 1; Length 257;  
Best Local Similarity 61.5%; Pred. No. 3.15e+01;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db f i k e  
Dt 363 ytyathargar 375  
OY 1 CTTCATTATGAA 13  
QT F I N E

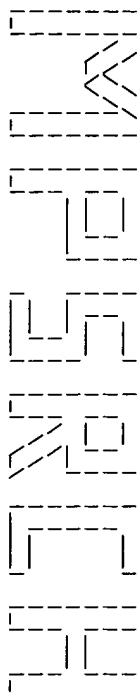
Search completed: Wed Nov 24 02:28:45 1999  
Job time : 3 secs.



THIS PAGE BLANK (0370)  
THIS PAGE BLANK (0370)

THIS PAGE BLANK (0370)



\*\*\*\*\*  
  
 (TM)  
 \*\*\*\*\*

Release 3.1a John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MParch\_nu n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 12:42:37 1999; MasPar time 1.71 Seconds  
 32.958 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-5  
 Description: (1-19) from US09103287.seq  
 Perfect Score: 19  
 N.A. Sequence: 1 CTTGATTAATGAGATGC 19  
 Comp: GAAGTATTACTTGCTACG

Scoring table: TABLE default  
 Gap 10

Match STD : Dbase 0; Query 0

Searched: 1052 segs, 1486975 bases x 2

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: HIV-NM8  
 1:H\_PRI 2:H\_UNA 3:H\_YIR

Statistics: Mean 6.893; Variance 1.353; scale 5.094

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	ID	Description	Pred. No.
C 1	63.2	672	3	SIV2010G	2.53e-01
C 2	63.2	781	3	HIV27924A1	2.53e-01
C 3	63.2	1773	3	SIVAGMT49	2.53e-01
C 4	63.2	9170	3	SIVAGMTYO	2.53e-01
C 5	63.2	9623	3	SIVAGMT677	2.53e-01
C 6	63.2	13068	3	SIVAGMT239	2.53e-01
C 7	57.9	672	3	SIV2010G	2.36e+00
C 8	57.9	781	3	HIV27924A1	2.36e+00
C 9	57.9	1217	1	HUMB2M1	2.36e+00
C 10	57.9	1773	3	SIVAGMT49	2.36e+00
C 11	57.9	8344	3	SIVAGMT49	2.36e+00
C 12	57.9	8344	3	EVAV	2.36e+00
C 13	57.9	9170	3	SIVAGMTYO	2.36e+00
C 14	57.9	9623	3	SIVAGMT677	2.36e+00
C 15	57.9	9625	2	SIVAGMT3	2.36e+00
C 16	52.6	270	3	HIVU08739	1.71e+01
C 17	52.6	275	3	HIVU08679	1.71e+01
C 18	52.6	275	3	HIVU08679	1.71e+01
C 19	52.6	276	3	HIVU08672	1.71e+01
C 20	52.6	276	3	HIVU08673	1.71e+01

21	10	52.6	276	3	HIVU08678	Human immunodeficiency	1.71e+01
22	10	52.6	279	3	HIVU08676	Human immunodeficiency	1.71e+01
C 23	10	52.6	781	3	HIV260415K	Human immunodeficiency	1.71e+01
C 24	10	52.6	1035	1	HUMEP1	Human immunodeficiency	1.71e+01
C 25	10	52.6	1113	3	SIVR3E	Human enhancer-binding	1.71e+01
C 26	10	52.6	1341	1	HUMTRP1	Human immunodeficiency	1.71e+01
C 27	10	52.6	1542	3	SIVAGM692	Human immunodeficiency	1.71e+01
C 28	10	52.6	1686	3	HIV2S52907	Human immunodeficiency	1.71e+01
C 29	10	52.6	1696	3	SIVAGMT17	Human immunodeficiency	1.71e+01
C 30	10	52.6	1715	3	HIVBR0141	Human immunodeficiency	1.71e+01
C 31	10	52.6	3201	3	HIVBAL2	Human immunodeficiency	1.71e+01
C 32	10	52.6	3534	3	SIMSP2	Simian spumavirus, pa	1.71e+01
C 33	10	52.6	3808	3	HIVBAL1	Human immunodeficiency	1.71e+01
C 34	10	52.6	8952	3	HIVL2	Human T-cell leukemia	1.71e+01
C 35	10	52.6	9215	2	SIVANDG1	Simian immunodeficiency	1.71e+01
C 36	10	52.6	9636	3	HIV2ISYR	Human immunodeficiency	1.71e+01
C 37	10	52.6	9671	3	HIV2ROD	Human immunodeficiency	1.71e+01
C 38	10	52.6	9672	2	HIV2ST	Human immunodeficiency	1.71e+01
C 39	10	52.6	9672	2	HIV2ST	Human immunodeficiency	1.71e+01
C 40	10	52.6	9811	3	SIVCP2	Simian immunodeficiency	1.71e+01
C 41	10	52.6	9811	3	SIVCP2	Simian immunodeficiency	1.71e+01
C 42	10	52.6	10271	3	HIV2UC1	Human immunodeficiency	1.71e+01
C 43	10	52.6	10359	2	HIV2BEN	Human immunodeficiency	1.71e+01
C 44	10	52.6	11443	3	HIV2CAM2	Human immunodeficiency	1.71e+01
C 45	10	52.6	11443	3	HIV2CAM2	Human immunodeficiency	1.71e+01

#### ALIGNMENTS

RESULT 1  
 LOCUS SIV2010G 672 bp DNA VRL 24-AUG-1994  
 DEFINITION Simian immunodeficiency virus (SIVbab) partial gag sequence,  
 yellow baboon sample SIVbab2010.  
 ACCESSION U10899  
 SOURCE Simian immunodeficiency virus from PCR-amplified DNA extracted from uncultured leukocytes derived from an asymptomatic pregnant female yellow baboon from Tanzania.  
 ORGANISM Simian immunodeficiency virus.  
 Vira; Viruses; ssRNA enveloped viruses; Retroviridae; Lentivirus; Primate immunodeficiency viruses.  
 1 (bases 1 to 672)

REFERENCE Jin, M.J., Rogers, J., Phillips-Conroy, J.E., Allan, J.S., Destrofers, R.C., Shaw, G.M., Sharp, P.M., and Kahn, B.H.  
 Infection of a yellow baboon with SIV from African green monkeys: evidence for cross-species transmission in the wild  
 J. Virol. (in press, 1994)  
 JOURNAL full staff review  
 STANDARD  
 COMMENT Two wild yellow baboons from Tanzania were identified in a previous study with strong SIV(gag) seroreactivity. They lived in a national park where yellow baboons and African green monkeys share the same habitat. In this study, molecular characterization was performed for virus derived from one of the SIV(gag) seroreactive yellow baboons (animal 2010) and from two vervet African green monkeys (animals ver266 and ver385). With this aim in mind, uncultured leukocyte DNA was PCR-amplified, cloned and sequenced over a 672 bp region of gag and a 906 bp region of env for animal 2010 (only the env region was analyzed for isolates ver266 and ver385). Jin et al. note that the viral strain derived from the yellow baboon (animal 2010) clusters with SIV strains of the veret subtype, including ver266 and ver385, and believe that this is the first evidence for simian-to-simian cross-species transmission of SIV in the wild.

#### FEATURES

##### source

##### CDS

1..672  
 /clone="gagbab2010"  
 /strain="SIVbab2010"  
 /organism="Simian immunodeficiency virus"  
 /specific\_host="Papio cynocephalus-hamdryas"  
 <1..>672  
 /gene="gag"  
 /codon\_start=1  
 /product="gag protein"  
 /translation="VATVQCCCHLVKESAAETSSGKROKNDRSTATSSGOSQNPFAQ"



OCGNTWVHVPLSPRTLNATKAVEKKFGAATVPMPOALSECTPDYDINOMLVGDH  
 OGALQIVKTEINEEAQMDIQPPRAGPFPACOLDRPSSDVGATSTTQOELENIYT  
 ANPRVDGVAIRRMWILGLOKCVKMNPNVILIDKOGKPEPKDYVDREYRTIRAEQA  
 SEVKKO"

BASE COUNT 245 a 141 c 161 g 125 t  
 ORIGIN approximately 304 bp downstream from the start of gag

Query Match 63.2%; Score 12; DB 3; Length 672;  
 Best Local Similarity 100.0%; Pred. No. 2.53e-01;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 335 tcaataatgaag 346  
 |||||||  
 Cp 12 TCATTATGAG 1

RESULT 2 HIV7924A1 781 bp ds-DNA VRL 23-AUG-1994  
 LOCUS Human immunodeficiency virus type 2 (HIV-2) p16/p28 gag sequence,  
 DEFINITION sample 7924A, clone 1, partial cds.  
 L33081  
 ACCESSION Human immunodeficiency virus type 2 PCR-amplified DNA recovered from  
 SOURCE a 48 year old heterosexual male originally from urban Guinea Bissau  
 Human immunodeficiency virus type 2  
 Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;  
 ORGANISM Retroviridae; Lentivirinae.  
 REFERENCE 1 (bases 1 to 781)  
 AUTHORS Gao,F., Yue,L., Robertson,D.L., Hill,S.C., Hul,H., Biggar,R.J.,  
 Nequave,A.E., Whelan,T.M., Ho,D.D., Shaw,G.M., Sharp,P.M. and  
 Hahn,B.H.  
 TITLE Genetic diversity of human immunodeficiency virus type 2: Evidence  
 for five distinct sequence subtypes with differences in virus  
 biology.  
 J. Virol. 68, 7433-7447 (1994)  
 JOURNAL  
 STANDARD  
 COMMENT Full staff review  
 This sequence was obtained as part of a study designed to determine  
 the in vivo extent of HIV-2 genetic and biological variation. Up  
 to now, HIV-2 sequences have been derived almost solely from  
 cultured isolates. Since cultivation may select against viruses  
 which can not grow in a particular cell type, Gao et al. argue that  
 a greater variation may be elucidated by sequencing virus obtained  
 from uncultured PBMCs. With this aim in mind, peripheral blood  
 samples were collected from 12 HIV-2 seropositive patients (2238,  
 60415K, 60667K, 7312A, 7810A, 7924A, FA, FO784, ET, UB, ON, and PA)  
 from six different West African countries, living in both urban and  
 rural areas. Uncultured PBMC DNA was PCR amplified, and the  
 recombinant clones were manually sequenced. Another portion of the  
 cells was used for virus isolation.  
 Patient 7924 was a heterosexual 48-year old teacher from Guinea  
 Bissau, who was living in the Washington, D.C. area. Originally he  
 lived in an urban area in West Africa before traveling extensively  
 and coming to the United States. After seeking medical care for a  
 STD, patient 7924 was identified as HIV-2 positive. Subsequent  
 Western blot analysis of peripheral blood cells showed HIV dual  
 reactivity. Virus isolation was successfully performed at the  
 University of Alabama at Birmingham. Sample 7924A clusters with  
 the subtype A HIV-2 viruses.  
 NCBI gi: 532145

FEATURES  
 source  
 Location/Qualifiers  
 1..781  
 /organism="Human immunodeficiency virus type 2"  
 /providal  
 /sequenced\_mol="DNA"  
 /tissue\_type="PBMC"  
 <1..>781  
 /partial  
 /gene="gag"  
 /note="NCBI gi: 532146"  
 /codon\_start=2  
 /translation="ANELDRFGLASLESKEGCGRIISVLDPLVPTGSENKLSFNT  
 VCVVCIHAEEKVDTEAKKIVORHVAETETAKMPTSPRTAPSGKGNYPVOS  
 IGGNTYHVPISPRITNATKAVEKKFGAATVPMPOALSECTPDYDINOMLVGDH

AAMQIRIRINEVADMDVAPRIPGAPLAPAGOLRPRSGDIAGTSTVEQIOWMFRAO  
 NPVPGNTYRIRWIOGLOKCVKMNPNVILIDKOGKPEPKDYVDREYRTIRAEQA  
 SEVKKO"

BASE COUNT 279 a 157 c 195 g 150 t  
 ORIGIN 111 bp downstream from the start of gag

Query Match 63.2%; Score 12; DB 3; Length 781;  
 Best Local Similarity 100.0%; Pred. No. 2.53e-01;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 507 tcaataatgaag 518  
 |||||||  
 Cp 12 TCATTATGAG 1

RESULT 3 SIVAGMT49 1773 bp ds-DNA VRL 09-JUN-1993  
 LOCUS Simian immunodeficiency virus, isolate TAN-49, complete gag cds.  
 DEFINITION L19253  
 ACCESSION Simian immunodeficiency virus proviral DNA, isolate TAN-49 from a  
 SOURCE Ugandan tantalus monkey (Cecopithecus tantalus).  
 ORGANISM Simian immunodeficiency virus  
 Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;  
 REFERENCE 1 (bases 1 to 1773)  
 AUTHORS Hirsch,V.M., McGonn,C., Dapolito,G., Goldstein,S., Ogen-Odoi,A.,  
 Biriyawho,B., Lakwo,T., and Johnson,P.R.  
 TITLE Identification of a new subgroup of SIVagm in tantalus monkeys  
 JOURNAL Virology (1993) In press  
 JOURNAL  
 STANDARD  
 COMMENT Kindly submitted prior to publication by Dr. Vanessa Hirsch,  
 Immunodeficiency Viruses Section, NIAID/NIH, Twinbrook II,  
 Rockville MD 20852. SIVagm-tan strains cluster with other  
 previously characterized agm strains taken from vervets and  
 gibbons. See related tantalus isolate sequences 17, 27, 40, and  
 9 (accession numbers L19250-L19252, L19254).  
 FEATURES  
 exon  
 Location/Qualifiers  
 195..1773  
 /gene="gag"  
 1..1773  
 /organism="Simian immunodeficiency virus"  
 /cell\_line="CEMS"  
 /haplotype="na"  
 /providal  
 /sequenced\_mol="DNA"  
 /codon\_start=2  
 /translation="SIVAGMT49 9170 bp ss-RNA  
 LOCUS SIVAGMT49 9170 bp ss-RNA VRL 01-MAR-1989  
 DEFINITION Simian (African green monkey) immunodeficiency virus, isolate  
 TYO-1, complete genome.  
 X07805  
 ACCESSION Simian (African green monkey) immunodeficiency virus, extra-  
 SOURCE chromosomal closed-circular DNA, TYO-1 clone lambda-SAH12.  
 REFERENCE 1 (bases 1 to 9170)  
 AUTHORS Fukasawa,M., Miura,T., Hasegawa,A., Morikawa,S., Tsujimoto,H.,  
 Miki,K., Kitamura,T. and Hayami,M.  
 TITLE Sequence of simian immunodeficiency virus from African green  
 JOURNAL monkey, a new member of HIV/STV group  
 STANDARD Nature 333, 457-461 (1988)  
 COMMENT This sequence was obtained from a lambda clone derived from an  
 African green monkey imported from Kenya.







```

exon          6135..6195
              /number=2
              /gene="rev"
exon          6202..8766
              /gene="env"
              /codon_start=6202
              8368..8449
              /number=3
              /gene="lat"
              8368..8750
              /number=3
              /gene="rev"
              8600..9271
              /gene="nef"
              /codon_start=8600
              8937..9623
              /note="3' long terminal repeat"
LTR           9407..9623
              /rpt_type=repeat 3' copy
              9506..9511
              /polyA_signal
BASE COUNT    3243 a 1807 c 2470 g 2103 t
ORIGIN        5' terminus of 5' LTR.

Query Match   63.2%   Score 12; DB 3; Length 9623;
Best Local Similarity 100.0%; Pred. NO. 2.53e-01;
Matches       12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db            1513 tcattaatgaag 1524
              |||||||
 Cp            12 TCATTATGAAG 1

```

RESULT	6	STIVMAC239	13068 bp ss-RNA	VRL	15-OCT-1991
LOCUS		Simian immunodeficiency virus isolated from a macaque.			
DEFINITION		Isolate 239; complete proviral genome and flanking sequence.			
ACCESSION		M33262 M61062-M61093			
SOURCE		Simian immunodeficiency virus (SIV) from a macaque, isolate 239 (Macaca mulatta Mm239-82); proviral DNA. Infectious clone, lambda stx239-1 [3]. Clone-infected peripheral blood lymphocytes from rhesus macaques [4]. Integrated proviral DNA in lambda vectors transfected into H9-78 cells [5].			
REFERENCE		1 (bases 1 to 10535) Regier,D.A. and Desrosiers,R.C. The complete nucleotide sequence of a pathogenic molecular clone of simian immunodeficiency virus			
AUTHORS		Simian immunodeficiency virus			
TITLE		AIDS Res. Hum. Retroviruses 6, 1221-1231 (1990)			
JOURNAL					
STANDARD					
REFERENCE		2 (bases 10536 to 13068) Regier,D.A. and Desrosiers,R.C. personal communication (12-91)			
AUTHORS		personal communication from D.A. Regier, Dept. of Microbiology,			
TITLE					
JOURNAL					
STANDARD					
REFERENCE		3 (bases 1 to 10535) Kestler,H., Kodama,T., Ringler,D., Martin,M., Pedersen,N., Lackner,A., Regier,D., Sehgal,P., Daniel,M., King,N. and Desrosiers,R. Induction of AIDS in Rhesus monkeys by molecularly cloned simian immunodeficiency virus			
AUTHORS		Science 248, 1109-1112 (1990)			
TITLE					
JOURNAL					
STANDARD					
REFERENCE		4 (bases 6860 to 8434) Burns,D.P.W. and Desrosiers,R.C. Selection of genetic variants of simian immunodeficiency virus in persistently infected rhesus monkeys			
AUTHORS		J. Virol. 65, 1843-1854 (1991)			
TITLE					
JOURNAL					
STANDARD					
REFERENCE		5 (bases 9026 to 9145) Naidu,Y.M., Kestler,H.W.III, Li,Y., Butler,C.V., Silva,D.P., Schmidt,D.K., Troup,C.D., Sehgal,P.K., Sonigo,P., Daniel,M.D. and Desrosiers,R.C. Characterization of infectious molecular clones of simian immunodeficiency virus (SIVmac) and human immunodeficiency virus			
AUTHORS					
TITLE					

JOURNAL	6 (sites; see <SIYMM1A11>)
STANDARD	Lutw,P.A., Shaw,K.E., Unger,R.E., Planellies,V., Stout,M.W.,
REFERENCE	Prittlow,E., Leung,N.J., Banapur,B. and Matthias,M.L.
AUTHORS	Genetic and Biologic Comparisons of Pathogenic and Non-pathogenic
TITLE	Molecular Clones of Simian Immunodeficiency Virus (SIVmac)
JOURNAL	AIDS Res. Hum. Retroviruses (1991) In press
STANDARD	7 (sites)
REFERENCE	Kestler,H.W., Ringler,J., Mori,K., Panicali,D.L., Sehgal,P.K.,
AUTHORS	Daniel,M.D. and Desrosiers,R.C.
TITLE	Importance of the nef gene for maintenance of high virus loads and
JOURNAL	for development of AIDS
STANDARD	Cell 65, 651-662 (1991)
COMMENT	[1, 2] kindly submitted in computer readable form prior to publication by Dr. Desrosiers, New England Regional Primate Center, Harvard. 256 bases of 5' flanking cellular DNA and approximately 2500 bases of 3' flanking material were also determined.
	[3] describes the pathogenic potential of SIVM239.
	27 late-time-point clones obtained from 2 rhesus monkeys infected with SIVM239 were used to quantify variation in the gp120 region of env [4].
	Macaque lymphocytes supported the replication of SIVM239 and SIVM251 but not SIVM142 or HIVZROD [5].
	See also <SIYMM1A11> (accession number M76764); SIVM239 and SIVM1A11 show greater than 98% homology at the nucleotide level. Analysis in [6] largely focuses on sequence differences in SIVM239 and SIVM1A11 which may account for differences in biologic properties [6].
	This entry replaces pages I-B-86 through I-B-89 (APR 90) in the 1990 HUMAN RETROVIRUSES AND AIDS.
FEATURES	Location/Oualifiers
exon	<6558..6853
	/note-"tat protein, exon 2 (first expressed exon)"
exon	<6784..6853
	/note-"rev protein, exon 2 (first expressed exon)"
CDS	join(6558..6853,9062..9158)
	/note-"tat protein"
	/codon_start=1
	/translation="METPLREQNSLESSNERSSCISEADASTPESANLGEIISQLY RPLEACNYTCCKKCCYCHQCFLKGLGICYEOSRRRRTPPKAKANTSSANKPIYS NRTRHCOPEKAKEIYEVAAVAATAGLR"
	join(6784..6853,9062..9315)
	/note-"rev protein"
CDS	/codon_start=1
	/translation="MSNHREEELRKRLRIHLHQTNPPYPTGGTANORQRARRMR RRMOQLALADRITYSPDPPTDLDAIOLOMLAIESIIPDTNTPEALCDPTEDSD RSPOD"
	1..256
cellular	/note-"flanking cellular DNA"
provirus	257..10535
	/note-"SIVmac239 provirus"
LTR	257..1074
	/note-"5'LTR"
LTR	257..773
	/note-"U3"
repeat_region	774..950
	/note-"R"
misc_feature	775
	/note-"mRNA start (putative)"
LTR	951..1074
	/note-"U5"
misc_binding	1078..1105



```
CDS
/note="tRNA primer binding site"
1309..2841
/note="gag polyprotein"
/codon_start=1
/translation="MGVRSVLSGKKKADLEKIRLRPNCKKKYMLKHVYMANELDRF
GLASLEENKRECCOKITSLYLAIPYSGSNLKLNTVCVIMCHIAEEVYKTEERAKO
LYORHLVETGETTEPTKTSRPTAPSSGGNGATPVQOIGENTVHPLSRITLNAVKL
IEERKGAENVFQALSECTPYDINQMLNCVGHQAQIIRDIINEEADMDLQ
POAPQOGLREPSGSDIAGTSSVDEQIOMNYROONPIPVNIYRMVQLQICVYR
MYNPNIIDKOGKPEPSYDREFKSLRABQTDAAVNMVOTILIONAPRDKLV
LKGIVNPTLEMLTACOGVGGPGQALMALALAPVLPFAAQQORREPRBK
CMNCKREGHSAROCRAPRQCGMKCKNDHVAKCPDRAGLGLGCPCKRRNPMMA
OVHGMPLTAPEDPAVDLKNITMOLGKQREKREKREKPEYTEDLLHLSLFGG
DQ"
2484..5666
/note="pol polyprotein (AA at 2484)"
/codon_start=1
/translation="MELMERITLCKAMSPKRTGMLKMKNGPCYCGMPRTGCFR
FMSGKEAPQPHSGSAGDANCSPRGSCSASLEHVAAGAAERKARKOREALQG
GDRGAAPQFSLMRBPVVAHLEGQVEVLDTGADDSIVTGEIDGPHPTPTVGGIG
GFINTREKYNVIEVLGKRIKGTIMTGTPIPIFGRNLTALGMLNPIAKVEPVY
ALKPGKDGKRLKMPLEKREKVALREICEMKEDQLEAPPTNPNPTFAIKKKD
NKMRLIDERELNRYTODETEVOGLIPHPAGLAKRRITVLIDGAYTSIPDEEPR
YTAFITPSYNNABPKRTIYKVLPOGKSPALFOYTMHVLPEPRKANPDVTLQYM
DITLASDRTDLEHRYVQSKELNSIGFSTPEEFQDPPQWNGELMPTKMLQ
KILPORERTVNDIOLKLVGNMAAQIPIGKTRHLCRLINGKMTLVEVQWTEMAE
AEEENKIIISOEGCYOEGKPLEATVKSQDNOMSYKIQEDKILKVGKFAIKN
HTNGVRLLAHYIOKIGEAIVMGQVPEFHLPEVDQWQWOTYMOVWTIPEMDFI
STPLVRLVLENTKDIPEGTEYVYDGSCKSKGKAGITDRGDKRKYKVEVQTNQ
QAELEFALMALDSSGRKANITVDSQYVAKITTCPTRESSRLVNOIEMIKSEIY
AAVPAHKGIGNQEIDHLVSQIRQVLELEKIEPAOEEDKTHSVKELVEFEKPRI
VARQIYVDCDKHQKGAHLGQNSDLGQMDCDHLBEKIIIVAVHVASGPIEAIRI
FOETROALFELKLAGWPITHLTDNANASQEVKVMAMAEIHEFGVGPYDOS
QGVVAMNHKLKQIDRIEQAQNSVETIYLVAMVHGMCFNRGSGIDMPFAELNMIT
TDEIOFOOSKSKFRKERNERYVREGDOLMKPGLMLKGBEAVILKVTGDIKVPFR
KAKITKDGKGEVDSSTIMEPTGEREVA"
5340..5355
/note="polypurine tract"
5340..5355
/note="vif protein"
/codon_start=1
/translation="MEERKMTAVPTMRPERLEHNSILKLYKKTDLQVCYVPH
KRVGAWMTCSRVITPLOGSHLEVQVGHILPEKGMSTAVRITWSKNWTVP
NADILHSTYPCETAGVRAIRGEQLSCCRPPRAKRYVPSLQVIALVADVR
SGENPTKQMDRNRGRIRMAKONSQDQGRKPRPTGANEPLGAVLGLILA"
5068..6406
/note="vpx protein"
/codon_start=1
/translation="MSDRERIPGNSGEETIGEAPEWLNRTVEEINREAVNHLPREL
IFQVQRSWEYHDEQMSPYKYRYLCLLOKALFMCKKCCRLGEGHGAGWRPG
PPPPPGCLA"
6407..6712
/note="vpr protein"
/codon_start=1
/translation="MEERPENEGQREPMDEWVEVLEELKEALKHEDPRLTALG
NHLYNRHGDTLGAAGELLRIIOLRALFMHRRGCSIRISLQPGGNDLSAIPSRSLM"
6860..9499
/note="env polyprotein"
/codon_start=1
/translation="MGCIAGNQLLAILLISVYIGYCLTVYVYGVAPMANATIDLCF
ARKNDTWGTCQCLPDNGDYSEVALNVTESFPAMNTVTEQAIEDVWOLFESISPCV
KLSPLCITWRCKNSSETDRLWGLSITTAISTASAVADVAMNETSSCIADQNDGL
BOEOMISCFENMTGLKDRKKEYNENYVADLVCOGNNTGESRCYMHNCQTSYIOE
SCDKHYMDAIRRYCARGVALLRNDNTVSGFMRCSKVYVSSCTRAMENQTSWFG
FNGTRAENRTIYMHGRDRTIISLKNYINLMKCRPCKNTVLYVTIISGLVFIHSOP
INDRPQAWCMFEGKWKDAIKEVKOTIVAHPRYTGINNTDKINTLAPGGGDEVFMM
TNGREFLYCKMNFELMWEDNRTNOKREKQKRYVCHIROIINMHKGVKAVYU
PREGDLICNSTVTSILANIDWDIGNQITNITSAEVALYRELGLYKLEITPIGLA
PVDVRYTGTGTSNNKRGVAVGLGELTAGSAMAASILTALQSRTLLAIIVOOQ
OLLDVYKROOELLRLTWGTKNLOTRTVTAIEKYLDOALNMWGAARVQCTGYTDP
NNSLTPKNNETWQEMERKVDPLEEITTLLEBAIQOEKKNTELOKLSMDYTFWME
DIASWTKIYQVYIVAGVILRLIYIVYOMLAKROGRPFSSPSTFOOTHQOD
PALPTREGKERDGGGNSWPMOIEYIHFILROLIRLLTLFNSCRTLLRSVYOIL
```

```
FEATURES
source
CDS
exon
exon
CDS
CDS
Query Match 63.28; Score 12; DB 3; Length 13068;
Best Local Similarity 92.98; Pred. No. 2,53e-01;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 203 gcatcttcattaa 216
Cp 19 GCATCCTTCAITTA 6
RESULT 7
LOCUS STV2010G 672 bp DNA VRL 24-AUG-1994
DEFINITION Simian immunodeficiency virus (STV2010) partial gag sequence,
YELLOW baboon sample STV2010.
ACCESSION U010899
SOURCE Simian immunodeficiency virus from PCR-amplified DNA extracted from
uncultured leukocytes derived from an asymptomatic pregnant female
yellow baboon from Tanzania.
ORGANISM Simian immunodeficiency virus
Virus; Viruses; ssRNA enveloped viruses; Retroviridae; Lentivirus;
Primate immunodeficiency viruses.
REFERENCE 1 (bases 1 to 672)
Jin,M.J., Rogers,J., Phillips-Conroy,J.E., Allan,J.S.,
Destrochers,R.C., Shaw,G.M., Sharp,P.M. and Hahn,B.H.
Infection of a yellow baboon with STV from African green monkeys:
evidence for cross-species transmission in the wild
J. Virol. (in press, 1994)
COMMENT Two wild yellow baboons from Tanzania were identified in a previous
study with strong STV(gag) seroreactivity. They lived in a
national park where yellow baboons and African green monkeys share
the same habitat. In this study, molecular characterization was
performed for virus derived from one of the STV(gag) seroreactive
yellow baboons (animal 2010) and from two vervet African green
monkeys (animals ver266 and ver385). With this aim in mind,
uncultured leukocyte DNA was PCR-amplified, cloned and sequenced
over a 672 bp region of gag and a 906 bp region of env for animal
2010 (only the env region was analyzed for isolates ver266 and
ver385). Jin et al. note that the viral strain derived from
the yellow baboon (animal 2010) clusters with STV strains of the
vervet subtype, including ver266 and ver385, and believe that
this is the first evidence for simian-to-simian cross-species
transmission of STV in the wild.
location/Qualifiers
1..672
/clone="gagbab2010"
/strain="STVbab2010"
/organism="Simian immunodeficiency virus"
/specific_host="Papio cynocephalus hamadryas"
<1..>672
/gene="gag"
/codon_start=1
/product="gag protein"
/translation="VATYRQCCHLVKKKSAETSSGCKNDRTATSSGSGQNFPAQ
QOQNTWVAVPSPRLNVAWRAVEKRGATVPMFQALSGCTPYDINQMLNVLGDH
OGALQIVEIINEEAQMDIOPHPAGLPGQALRPGVAGTSTVQDLEWITV
ANPRVDGAIYRWRITLQRCVKNVNSIILDKQKPEFKYDVDRFYRTIRAEQA
```



BASE COUNT 245 a 141 c 161 g 125 t  
 ORIGIN approximately 304 bp downstream from the start of gag

Query Match 57.9%; Score 11; DB 3; Length 672;  
 Best Local Similarity 100.0%; Pred. No. 2.36e+00;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 335 tcattatgaa 345  
 |||||||  
 3 TCATTATGAA 13

RESULT 8 HIV27924A1 781 bp ds-DNA VRL 23-AUG-1994  
 LOCUS Human immunodeficiency virus type 2 (HIV-2) p16/p28 gag sequence,  
 DEFINITION sample 7924A, clone 1, partial cds.  
 L33081

ACCESSION Human immunodeficiency virus type 2 PCR-amplified DNA recovered from  
 SOURCE a 48 year old heterosexual male originally from urban Guinea Bissau  
 who showed HIV dual seroreactivity.  
 Human immunodeficiency virus type 2  
 Viridae: ss-RNA enveloped viruses; Positive strand RNA virus;  
 Retroviridae: Lentivirinae.  
 1 (bases 1 to 781)  
 Gao,F., Yue,L., Robertson,D.L., Hill,S.C., Hu,I., Biggar,R.J.,  
 Neequaye,A.E., Whelan,T.M., Ho,D.D., Shaw,G.M., Sharp,P.M. and  
 Hahn,B.H.

REFERENCE Genetic diversity of human immunodeficiency virus type 2: Evidence  
 AUTHORS for five distinct sequence subtypes with differences in virus  
 biology  
 J. Virol. 68, 7433-7447 (1994)

STANDARD full staff review  
 COMMENT This sequence was obtained as part of a study designed to determine  
 the in vivo extent of HIV-2 genetic and biological variation. Up  
 to now, HIV-2 sequences have been derived almost solely from  
 cultured isolates. Since cultivation may select against viruses  
 which can not grow in a particular cell type, Gao et al. argue that  
 a greater variation may be elucidated by sequencing virus obtained  
 from uncultured PBMCs. With this aim in mind, peripheral blood  
 samples were collected from 12 HIV-2 seropositive patients (2238,  
 60415K, 60667K, 7312A, 7810A, 7924A, FA, F0784, FT, JA, ON, and PA)  
 from six different West African countries, living in both urban and  
 rural areas. Uncultured PBMC DNA was PCR amplified, and the  
 recombinant clones were manually sequenced. Another portion of the  
 cells was used for virus isolation.  
 Patient 7924 was a heterosexual 48-year old teacher from Guinea  
 Bissau, who was living in the Washington, D.C. area. Originally he  
 lived in an urban area in West Africa before traveling extensively  
 and coming to the United States. After seeking medical care for a  
 STD, patient 7924 was identified as HIV-2 positive. Subsequent  
 Western blot analysis of peripheral blood cells showed HIV dual  
 reactivity. Virus isolation was successfully performed at the  
 University of Alabama at Birmingham. Sample 7924A clusters with  
 the subtype A HIV-2 viruses.  
 NCBI 91: 532145

FEATURES  
 source Location/Qualifiers  
 1..781  
 /organism="Human immunodeficiency virus type 2"  
 /proviral  
 /sequenced\_mol="DNA"  
 /tissue\_type="PBMC"  
 <1..>781  
 /partial  
 /gene="gag"  
 /note="NCBI 91: 532146"  
 /codon\_start=2  
 /translation="ANELDRFLAESLSEKGGORIIIVLDPIVPGSENLKSLFNT  
 VCVVMIHAEEKVDTEAKRIVORHVAETGTAEKMPNTRPAPSGKGNPVVS  
 IGVNTHVPLSPRTLNAMVKVEEKFGAALVEGCTGPDINDNOMNCVGDHO  
 AAMGIRIINEVADMDVARIPIGRLPAGLRRPGRSDINGTSTYEDQIQMFRRA  
 NPVVGNIYRIRIOTIGLOKCYRMRNPTIILDVYKGGPRSFOS"

BASE COUNT 279 a 157 c 195 g 150 t

ORIGIN 111 bp downstream from the start of gag

Query Match 57.9%; Score 11; DB 3; Length 781;  
 Best Local Similarity 100.0%; Pred. No. 2.36e+00;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 507 tcattatgaa 517  
 |||||||  
 3 TCATTATGAA 13

RESULT 9 HUMB2M1 1217 bp ds-DNA PRI 15-JUN-1989  
 LOCUS Human beta-2-microglobulin gene, exon 1.  
 DEFINITION M17986  
 ACCESSION 1 of 2  
 SEGMENT  
 SOURCE Human DNA, clone pb2m13, and cDNA to mRNA, clone pb2m16.  
 ORGANISM Homo sapiens  
 Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;  
 Eutheria; Primates; Haplorhini; Catarrhini; Homidae.

REFERENCE 1 (bases 1 to 1217)  
 Kottman,D., Rein,R., Ginjar,I., Hochstenbach,F., Seemann,G.,  
 The human beta-2-microglobulin gene: Primary structure and  
 definition of the transcriptional unit  
 J. Immunol. 139, 3132-3138 (1987)

AUTHORS  
 TITLE Full staff review  
 STANDARD Draft entry and computer-readable sequence for [1] kindly provided  
 COMMENT by E.J.Baas, 03-DEC-1987.  
 FEATURES  
 Location/Qualifiers  
 exon  
 <883..949  
 /number=1  
 /note="beta-2-microglobulin precursor"  
 /gene="B2M"  
 /map="15q21-q22.2"  
 prim\_transcript 846..>1217  
 /note="microglobulin mRNA and intron"  
 sig\_peptide 883..942  
 /note="beta-2-microglobulin, signal peptide"  
 mat\_peptide 943..949  
 /codon\_start=1  
 /note="beta-2-microglobulin, mature peptide"  
 intron 950..>1217  
 /codon\_start=1  
 /note="intron A"  
 source 1..1217  
 /organism="Homo sapiens"  
 BASE COUNT 274 a 325 c 322 g 296 t  
 ORIGIN 383 bp upstream of EcoRI site.

Query Match 57.9%; Score 11; DB 1; Length 1217;  
 Best Local Similarity 100.0%; Pred. No. 2.36e+00;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 32 catcgatcat 42  
 |||||||  
 Cp 18 CATCGTCATT 8

RESULT 10 SIYAGT49 1773 bp ds-DNA VRL 09-JUN-1993  
 LOCUS Simian immunodeficiency virus, isolate TAN-49, complete gag cds.  
 DEFINITION I19253  
 ACCESSION  
 SOURCE Simian immunodeficiency virus proviral DNA, isolate TAN-49 from a  
 Ugandan tantalus monkey (Cecopithecus tantalus).  
 ORGANISM Simian immunodeficiency virus  
 Viridae: ss-RNA enveloped viruses; Positive strand RNA virus;  
 Retroviridae: Lentivirinae.  
 1 (bases 1 to 1773)  
 Hirsch,V.M., McCann,C., Dapolito,G., Goldstein,S., Ogen-Odoi,A.,  
 Biraywaho,B., Lakwo,T. and Johnson,P.R.  
 Identification of a new subgroup of SIYagm in tantalus monkeys  
 JOURNAL Virology (1993) in press



STANDARD full staff\_review  
 COMMENT kindly submitted prior to publication by Dr. Vanessa Hirsch, Immunodeficiency Viruses Section, NIAID/NIH, Twindbrook II, Rockville MD 20852. SIVagm-tan strains cluster with other previously characterized agm strains taken from vervets and grivets. See related tantalus isolate sequences 17, 27, 40, and 9 (accession numbers L19250-L19252, L19254).  
 FEATURES  
 exon Location/Qualifiers  
 source  
 1..1773  
 /gene="gag"  
 /organism="Simian immunodeficiency virus"  
 /cell\_line="CEMss"  
 /haplotype="na"  
 /proviral  
 /sequenced\_mol="DNA"  
 BASE COUNT 594 a 375 c 480 g 324 t  
 ORIGIN 193 nt upstream from the gag cds start.  
 Query Match 57.9%; Score 11; DB 3; Length 1773;  
 Best Local Similarity 100.0%; Pred. No. 2.36e+00;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 828 tcatatgaa 838  
 |||||||||  
 Oy 3 TCATTAATGAA 13

RESULT 11  
 LOCUS E1AV 8344 bp ss-RNA VRL 17-MAR-1989  
 DEFINITION Equine infectious anemia virus proviral DNA, complete genome.  
 ACCESSION M16575 M1137 K03334 M14855  
 SOURCE Equine infectious anemia virus (E1AV) proviral DNA (clones 1369 and 409-2 [1]).  
 REFERENCE 1 (bases 1 to 8229)  
 AUTHORS Kawakami,T., Sherman,L., Dahlberg,J., Gazit,A., Yaniv,A., Tronick,S.R. and Aaronson,S.A.  
 TITLE Nucleotide sequence analysis of equine infectious anemia virus proviral DNA  
 JOURNAL Virology 158, 300-312 (1987)  
 STANDARD full staff\_review  
 REFERENCE 2 (bases 324 to 5122)  
 AUTHORS Stephens,R.M., Casey,J.W. and Rice,N.R.  
 TITLE Equine infectious anemia virus gag and pol genes: Relatedness to viena and AIDS virus  
 JOURNAL Science 231, 589-594 (1986)  
 STANDARD full staff\_review  
 REFERENCE 3 (bases 2300 to 2950)  
 AUTHORS Chiu,I.-M.  
 TITLE Unpublished (1985) Meloy Labs Inc, Springfield, VA 22151  
 STANDARD full staff\_review  
 REFERENCE 4 (sites: pept)  
 AUTHORS Chiu,I.-M., Yaniv,A., Dahlberg,J.E., Gazit,A., Skuatz,S.F., Tronick,S.R. and Aaronson,A.  
 TITLE Nucleotide sequence evidence for relationship of AIDS retrovirus to lentiviruses  
 JOURNAL Nature 317, 366-368 (1985)  
 STANDARD full staff\_review  
 REFERENCE 5 (bases 4769 to 8048; 8069 to 8344)  
 AUTHORS Payne,S.L.  
 TITLE Unpublished (1987) Dept Biochem, Louisiana St Univ, Baton Rouge, LA  
 STANDARD full staff\_review  
 REFERENCE 6 (bases 4769 to 8048; 8069 to 8228)  
 AUTHORS Rushlow,K., Olsen,K., Stiegler,G., Payne,S.L., Montelaro,R.C. and Isel,C.J.  
 TITLE Lentivirus genomic organization: The complete nucleotide sequence of the env gene region of equine infectious anemia virus  
 JOURNAL Virology 155, 309-321 (1986)  
 STANDARD full staff\_review  
 REFERENCE 7 (bases 1 to 8229; correction of 5348 in [1] above)  
 AUTHORS Tronick,S.R. and Fuller,F.  
 TITLE Unpublished, personal communication

STANDARD full staff\_review  
 COMMENT Correction of [1] by [7] leads to agreement with [6] regarding the env cds. The sequence starting with coordinate 8230 is horse cellular DNA.  
 FEATURES  
 LTR Location/Qualifiers  
 1..321  
 /note="5' LTR"  
 repeat\_region 208..>208  
 /note="R repeat 5' copy"  
 misc\_binding 325..342  
 /note="primer (lys-tRNA binding site)"  
 variation 342..342  
 /note="c in [Virology 158, 300-312 (1987)], t in [Science 231, 589-594 (1986)]"  
 CDS 465..1925  
 /note="gag polypeptide (AA at 465)  
 /codon\_start=465  
 <1682..5122  
 /partial  
 /note="pol polypeptide (AA at 1682; NH2-terminus uncertain)"  
 variation 2341..2341  
 /note="t in [Virology 158, 300-312 (1987)], c in [unpublished (1985) Meloy Labs Inc, Springfield, VA 22151]"  
 variation 2349..2349  
 /note="t in [Virology 158, 300-312 (1987)], c in [Science 231, 589-594 (1986)]"  
 variation 2436..2436  
 /note="g in [Virology 158, 300-312 (1987)], a in [unpublished (1985) Meloy Labs Inc, Springfield, VA 22151]"  
 variation 2677..2677  
 /note="t in [Virology 158, 300-312 (1987)], a in [Science 231, 589-594 (1986)]"  
 variation 2865..2870  
 /note="ttggaag in [Virology 158, 300-312 (1987)], agtaa in [Science 231, 589-594 (1986)]"  
 variation 2870..2870  
 /note="g in [Virology 158, 300-312 (1987)], a in [unpublished (1985) Meloy Labs Inc, Springfield, VA 22151]"  
 variation 3777..3777  
 /note="g in [Virology 158, 300-312 (1987)], a in [Science 231, 589-594 (1986)]"  
 variation 3858..3858  
 /note="g in [Virology 158, 300-312 (1987)], a in [Science 231, 589-594 (1986)]"  
 variation 4529..4530  
 /note="ac in [Virology 158, 300-312 (1987)], ca in [Science 231, 589-594 (1986)]"  
 variation 4573..4573  
 /note="a in [Virology 158, 300-312 (1987)], g in [Science 231, 589-594 (1986)]"  
 CDS 5312..7891  
 /note="envelope polypeptide (AA at 5312)  
 /codon\_start=5312  
 variation 6348..6348  
 /note="a in [Virology 158, 300-312 (1987)], g in [Virology 155, 309-321 (1986)], [unpublished (1987) Dept Biochem, Louisiana St Univ, Baton Rouge, LA]"  
 variation 6350..6350  
 /note="g in [Virology 158, 300-312 (1987)], a in [Virology 155, 309-321 (1986)], [unpublished (1987) Dept Biochem, Louisiana St Univ, Baton Rouge, LA]"  
 variation 7060..7060  
 /note="c in [Virology 158, 300-312 (1987)], t in [Virology 155, 309-321 (1986)], [unpublished (1987) Dept Biochem, Louisiana St Univ, Baton Rouge, LA]"  
 CDS <7234..7641  
 /partial



```

/feature="3' ORF protein (AA at 7233: NH2-terminus
uncertain)"
/codon_start=7235
7909..8229
/feature="3' LTR"
8048..8069
/feature="gagtcgcattgtgacgct in [Virology 158, 300-312
(1987)], g in [Virology 155, 309-321 (1986)], [unpublished
(1987)] Dept Biochem, Louisiana St Univ, Baton Rouge, LA]"
8073..8073
/feature="g in [Virology 158, 300-312 (1987)], c in [Virology
155, 309-321 (1986)], [unpublished (1987)] Dept Biochem,
Louisiana St Univ, Baton Rouge, LA]"
8098..8098
/feature="a in [Virology 158, 300-312 (1987)], g in [Virology
155, 309-321 (1986)], [unpublished (1987)] Dept Biochem,
Louisiana St Univ, Baton Rouge, LA]"
8104..8104
/feature="g in [Virology 158, 300-312 (1987)], a in [Virology
155, 309-321 (1986)], [unpublished (1987)] Dept Biochem,
Louisiana St Univ, Baton Rouge, LA]"
8116..8193
/feature="R repeat 3' copy"
8118..8119
/feature="ga in [Virology 158, 300-312 (1987)], gca in
[Virology 155, 309-321 (1986)], [unpublished (1987)] Dept
Biochem, Louisiana St Univ, Baton Rouge, LA]"
2984 a 1358 c 1846 g 2156 t

BASE COUNT      2984 a 1358 c 1846 g 2156 t
ORIGIN

Query Match      57.9% Score 11; DB 3; Length 8344;
Best Local Similarity 92.3%; Pred. No. 2.36e+00;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 5202 cttcataagaa 5214
|||||
1 CTTCATATACAA 13

RESULT 12
LOCUS      EIAV      8344 bp ss-RNA      VRL      17-MAR-1989
DEFINITION Equine infectious anemia virus proviral DNA, complete genome.
ACCESSION M16575 M1337 K03334 M14855
SOURCE      Equine infectious anemia virus (EIAV) proviral DNA (clones 1369 and
409-2 [1]).
REFERENCE 1 (bases 1 to 8229)
AUTHORS      Kawakami,T., Sherman,L., Dahlberg,J., Gazit,A., Yaniv,A.,
Tronick,S.R. and Aaronson,S.A.
TITLE      Nucleotide sequence analysis of equine infectious anemia virus
proviral DNA
JOURNAL      Virology 158, 300-312 (1987)
STANDARD full staff_review
AUTHORS      2 (bases 324 to 5122)
AUTHORS      Stephens,R.M., Casey,J.W. and Rice,N.R.
TITLE      Equine infectious anemia virus gag and pol genes: Relatedness to
viana and AIDS virus
JOURNAL      Science 231, 589-594 (1986)
STANDARD full staff_review
AUTHORS      3 (bases 2300 to 2950)
AUTHORS      Chiu,I.-M.
TITLE      Unpublished (1985) Meloy Labs Inc, Springfield, VA 22151
JOURNAL      full staff_review
AUTHORS      4 (sites: pep)
AUTHORS      Chiu,I.-M., Yaniv,A., Dahlberg,J.E., Gazit,A., Skuatz,S.F.,
Tronick,S.R. and Aaronson,A.
TITLE      Nucleotide sequence evidence for relationship of AIDS retrovirus to
lentiviruses
JOURNAL      Nature 317, 366-368 (1985)
STANDARD full staff_review
AUTHORS      5 (bases 4769 to 8048; 8069 to 8344)
AUTHORS      Payne,S.L.
JOURNAL      Unpublished (1987) Dept Biochem, Louisiana St Univ, Baton Rouge, LA
STANDARD full staff_review
REFERENCE 6 (bases 4769 to 8048; 8069 to 8228)
AUTHORS      Rushlow,K., Olsen,K., Stiegler,G., Payne,S.L., Montelaro,R.C. and
Issel,C.J.
TITLE      Lentivirus genomic organization: The complete nucleotide sequence
of the env gene region of equine infectious anemia virus
JOURNAL      Virology 155, 309-321 (1986)
STANDARD full staff_review
AUTHORS      7 (bases 1 to 8229; correction of 5348 in [1] above)
AUTHORS      Tronick,S.R. and Fuller,F.
TITLE      Unpublished, personal communication
JOURNAL      full staff_review
STANDARD full staff_review
COMMENT      Correction of [1] by [7] leads to agreement with [6] regarding the
env cds. The sequence starting w/lyh coordinate 8230 is horse
cellular DNA.
FEATURES
LTR      1..321
/feature="5' LTR"
208..>208
/feature="R repeat 5' copy"
325..342
/feature="primer (Lys-tRNA binding site)"
342..342
/feature="c in [Virology 158, 300-312 (1987)], t in [Science
231, 589-594 (1986)]"
465..1925
/feature="gag polypeptide (AA at 465)
/codon_start=465
<1682..5122
/partial
/feature="pol polypeptide (AA at 1682; NH2-terminus
uncertain)"
/codon_start=1682
2341..2341
/feature="t in [Virology 158, 300-312 (1987)], c in
[unpublished (1985) Meloy Labs Inc, Springfield, VA
22151]"
2349..2349
/feature="t in [Virology 158, 300-312 (1987)], c in [Science
231, 589-594 (1986)]"
2436..2436
/feature="g in [Virology 158, 300-312 (1987)], a in
[unpublished (1985) Meloy Labs Inc, Springfield, VA
22151]"
2677..2677
/feature="t in [Virology 158, 300-312 (1987)], a in [Science
231, 589-594 (1986)]"
2865..2870
/feature="gag in [Virology 158, 300-312 (1987)], agtaaa in
[Science 231, 589-594 (1986)]"
2870..2870
/feature="g in [Virology 158, 300-312 (1987)], a in
[unpublished (1985) Meloy Labs Inc, Springfield, VA
22151]"
3777..3777
/feature="g in [Virology 158, 300-312 (1987)], a in [Science
231, 589-594 (1986)]"
3858..3858
/feature="g in [Virology 158, 300-312 (1987)], a in [Science
231, 589-594 (1986)]"
4529..4530
/feature="ac in [Virology 158, 300-312 (1987)], ca in
[Science 231, 589-594 (1986)]"
4573..4573
/feature="a in [Virology 158, 300-312 (1987)], g in [Science
231, 589-594 (1986)]"
5312..7891
/feature="envelope polypeptide (AA at 5312)
/codon_start=5312
6348..6348
/feature="a in [Virology 158, 300-312 (1987)], g in [Virology
155, 309-321 (1986)], [unpublished (1987)] Dept Biochem,

```



variation Louisiana St Univ, Baton Rouge, LA"

6350..6350

/note="g in [Virology 158, 300-312 (1987)], a in [Virology 155, 309-321 (1986)], [unpublished (1987)] Dept Blochem, Louisiana St Univ, Baton Rouge, LA]"

variation 7060..7060

/note="c in [Virology 158, 300-312 (1987)], t in [Virology 155, 309-321 (1986)], [unpublished (1987)] Dept Blochem, Louisiana St Univ, Baton Rouge, LA]"

<7234..7641

/partial

/note="3' ORF protein (AA at 7233; NH2-terminus uncertain)"

/codon\_start=7235

7909..8229

/note="3' LTR"

8048..8069

/note="gagtcgcgatttgagcgcgt in [Virology 158, 300-312 (1987)], gc in [Virology 155, 309-321 (1986)], [unpublished (1987)] Dept Blochem, Louisiana St Univ, Baton Rouge, LA]"

variation 8073..8073

/note="g in [Virology 158, 300-312 (1987)], c in [Virology 155, 309-321 (1986)], [unpublished (1987)] Dept Blochem, Louisiana St Univ, Baton Rouge, LA]"

variation 8098..8098

/note="a in [Virology 158, 300-312 (1987)], g in [Virology 155, 309-321 (1986)], [unpublished (1987)] Dept Blochem, Louisiana St Univ, Baton Rouge, LA]"

variation 8104..8104

/note="g in [Virology 158, 300-312 (1987)], a in [Virology 155, 309-321 (1986)], [unpublished (1987)] Dept Blochem, Louisiana St Univ, Baton Rouge, LA]"

repeat\_region 8116..8193

/note="R repeat 3' copy"

8118..8119

/note="ga in [Virology 158, 300-312 (1987)], gca in [Virology 155, 309-321 (1986)], [unpublished (1987)] Dept Blochem, Louisiana St Univ, Baton Rouge, LA]"

BASE COUNT 2984 a 1358 c 1846 g 2156 t

ORIGIN

Query Match 57.9%; Score 11; DB 3; Length 8344;

Best Local Similarity 92.3%; Pred. No. 2.36e+00;

Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 5203 ttccatgaagaag 5215

|||||

CP 13 TTCATTATGAG 1

RESULT 13

LOCUS SIVAGMTO 9170 bp ss-RNA VRL 01-MAR-1989

DEFINITION Simian (African green monkey) immunodeficiency virus, isolate TYO-1, complete genome.

ACCESSION X07805

SOURCE Simian (African green monkey) immunodeficiency virus, extra-chromosomal closed-circular DNA, TYO-1 clone lambda-SAH12.

REFERENCE 1 (bases 1 to 9170)

AUTHORS Fukusawa, M., Mura, T., Hasegawa, A., Morikawa, S., Tsujimoto, H., Miki, K., Kitamura, T. and Hayami, M.

TITLE Sequence of simian immunodeficiency virus from African green monkey, a new member of HIV/SIV group

JOURNAL Nature 333, 457-461 (1988)

STANDARD full staff\_review

COMMENT This sequence was obtained from a lambda clone derived from an African green monkey imported from Kenya.

SIVAGMTO appears to lack a vpr coding region. [1] suggests that this may account for SIVAGMTO's lack of pathogenicity. The env cds contains an in-frame stop codon at positions 8077-8079, similar to the in-frame stop codons found in the SIV macaque and HIV-2 RD35 envelope coding regions, except that SIVAGMTO's stop codon

is found 96bp downstream from the 3' splice junction of tat and rev, which is the location of the stop codon in the other sequences.

[1] finds that SIVAGMTO is approximately equally distantly related to the HIV-1's and the HIV-2/SIV (macaque) group; over a "conserved" stretch of the pol protein, TYO is only 87% similar to AGM385 and 82% similar to AGM266, the latter two differing by 20%.

FEATURES

LTR

CDS

<1634..4819

/note="5' LTR"

1634"

/note="pol polypeptide (NH2-terminus uncertain; AA at 1634)"

/codon\_start=1634

join(5546..5763,7984..8068)

/note="tat protein"

/codon\_start=5546

join(5703..5763,7984..8177)

/note="rev protein"

/codon\_start=5703

prim\_transcript 1..9170

/note="genomic mRNA"

prim\_transcript 1..9170

/note="tat, rev, nef subgenomic mRNA"

repeat\_region 1..117

/note="R repeat 5' copy"

220..237

/note="primer (Lys-tRNA) binding site"

432..1991

/note="gag polypeptide"

/codon\_start=432

4755..5462

/note="vif protein"

/codon\_start=4755

5236..5595

/note="vpx protein"

/codon\_start=5236

5546..5763

/note="tat protein, exon 2 (first expressed exon)"

5703..5763

/note="rev protein, exon 2 (first expressed exon)"

5764..7983

/note="tat intron 2"

5764..7983

/note="rev intron 2"

5764..7983

/note="tat, rev, nef subgenomic mRNA intron 2"

5770..8367

/pseudo

/note="env polypeptide (in-frame stop codon at 8077)"

/codon\_start=5770

7984..8068

/note="tat protein, exon 3 (AA at 7985)"

7984..8177

/note="rev protein, exon 3 (AA at 7986)"

8077..8079

/note="in-frame stop in env cds"

8201..8890

/note="nef protein"

/codon\_start=8201

8547..>9170

/note="3' LTR"

9054..9170

/note="R repeat 3' copy"

9151..9156

/note="mRNA polyadenylation signal"

BASE COUNT 3172 a 1720 c 2263 g 2015 t

ORIGIN

Cap site of genomic RNA.

Query Match 57.9%; Score 11; DB 3; Length 9170;

Best Local Similarity 100.0%; Pred. No. 2.36e+00;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Db 1066 tcattaatgaa 1076  
 |||  
 QY 3 TCATTATGAA 13

RESULT 14  
 LOCUS SIVAGM677 9623 bp ss-DNA VRL 02-APR-1991  
 DEFINITION Simian immunodeficiency virus from African Green Monkey;  
 gri-1 lambdaII or 677 clone (from a grivet); complete genome.  
 M66437 M29973  
 SOURCE Simian immunodeficiency virus from an African grivet (green monkey), gri-1, lambda clone II (biologically active clone); proviral DNA, also denoted 677.  
 1 (bases 1 to 9623)  
 Fomsgaard,A., Hirsch,V.M., Allan,J.S. and Johnson,P.R.  
 A highly divergent proviral DNA clone of SIV from a distinct species of African green monkey.  
 Virology 182,397-402 (1991)  
 JOURNAL full staff-entry  
 STANDARD 2 (bases 1 to 2438)  
 REFERENCE Johnson,P.R., Fomsgaard,A., Allan,J., Gravel,M., London,M.T., Olmstead,R.A. and Hirsch,V.M.  
 Simian immunodeficiency viruses from African green monkeys display unusual genetic diversity  
 J. Virol. 64, 1086-1092 (1990)  
 TITLE Simian immunodeficiency viruses from African green monkeys display unusual genetic diversity  
 JOURNAL full staff-entry  
 STANDARD full submitted prior to publication and in a computer readable form by Phillip Johnson, NIAID/Georgetown University.  
 COMMENT The typical stop codon "tag" in rev (position 8574) is "cag" making rev in SIVgri-1 (lambdaII) longer than rev in most other sequences. The gri-1 sequence is the most divergent primate lentivirus reported to date.  
 Location/Qualifiers  
 join(5993..6195,8370..8750)  
 /gene="tat"  
 /codon\_start=5993  
 /partial  
 join(6135..6195,8371..8750)  
 /note="The normal stop codon 'tag' in rev (position 8574) is 'cag' making rev in SIVgri-1 longer than rev in most other sequences. The gri-1 sequence is the most divergent primate lentivirus reported to date."  
 /gene="rev"  
 /codon\_start=6135  
 1..689  
 /note="5' long terminal repeat"  
 1..587  
 /rpt\_type=R repeat 5' copy  
 690..707  
 /note="primer (Lys-trNA) binding site"  
 897..2438  
 /gene="gag"  
 /codon\_start=897  
 2141..5314  
 /gene="pol"  
 /codon\_start=2141  
 /partial  
 5214..5873  
 /gene="vif"  
 /codon\_start=5214  
 5683..6039  
 /gene="vpr"  
 /codon\_start=5683  
 5993..6195  
 /number=2  
 /gene="tat"  
 6135..6195  
 /number=2  
 /gene="rev"  
 6202..8766  
 /gene="env"  
 /codon\_start=6202

exon 8368..8449  
 /number=3  
 /gene="tat"  
 8368..8750  
 /number=3  
 /gene="rev"  
 8600..9271  
 /gene="nef"  
 /codon\_start=8600  
 8937..9623  
 /note="3' long terminal repeat"  
 9407..9623  
 /rpt\_type=R repeat 3' copy  
 9506..9511  
 polyA\_signal  
 BASE COUNT 3243 a 1807 c 2470 g 2103 t  
 ORIGIN 5' terminus of 5' LTR.  
 Query Match 57.9%; Score 11; DB 3; Length 9623;  
 Best local similarity 100.0%; Pred. No. 2.36e+00;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1513 tcattaatgaa 1523  
 |||  
 QY 3 TCATTATGAA 13

RESULT 15  
 LOCUS SIVAGM3 9625 bp ds-RNA UNA 04-JAN-1990  
 DEFINITION Simian immunodeficiency virus from African green monkey,  
 isolate AGM3, complete genome.  
 M30931  
 SOURCE Simian immunodeficiency virus from African green monkey, isolate 3,  
 complete proviral genomic DNA. Infectious clone.  
 1 (bases 1 to 9625)  
 Baier,M., Garber,C., Mueller,C., Cichurek,K. and Kurth,R.  
 Complete nucleotide sequence of a simian immunodeficiency virus  
 from African green monkeys: a novel type of intragroup divergence  
 unpublished (1990); Paul-Ehrlich-Institute, Paul-Ehrlich-Str-57-59,  
 6070 Langen 1, Federal Republic of Germany  
 JOURNAL full staff-review  
 STANDARD  
 COMMENT Draft entry and computer-readable sequence for [1] kindly submitted  
 by M.Baier 20-DEC-1989.

FEATURES  
 CDS  
 location/Qualifiers  
 join(5554..5771,8013..8154)  
 /note="tat protein"  
 /codon\_start=5554  
 join(5711..5771,8013..8221)  
 /note="rev protein"  
 /codon\_start=5711  
 431..1996  
 /note="gag polyprotein"  
 /codon\_start=431  
 <1687..4827  
 /partial  
 /note="pol polyprotein (NH2 terminus uncertain: AA at 1687)"  
 /codon\_start=1687  
 4763..5461  
 /note="vif protein"  
 /codon\_start=4763  
 5244..5603  
 /note="vpx protein"  
 /codon\_start=5244  
 5554..5771  
 /note="tat protein, exon 2 (first expressed exon)"  
 5711..5771  
 /note="rev protein, exon 2 (first expressed exon)"

This sequence was taken from an infectious molecular clone (used for heterologous infection of the pigtail macaque). The 3' LTR sequence does not appear to match the 5' LTR sequence.



CDS 5778..8411  
 /note="env polypeptide"  
 /codon\_start=5778  
 exon 8013..8154  
 /note="tat protein, exon 3 (AA at 8014)"  
 exon 8013..8221  
 /note="rev protein, exon 3 (AA at 8015)"  
 CDS 8245..8934  
 /note="nef protein"  
 /codon\_start=8245  
 BASE COUNT 3324 a 1827 c 2383 g 2091 t  
 ORIGIN

Query Match 57.98; Score 11; DB 2; Length 9625;  
 Best Local Similarity 86.7%; Pred. No. 2.36e+00;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Db 7139 atctgcatatga 7153  
 ||| |||||  
 CP 17 ATGCTCATTAATGA 3

Search completed: Sat Nov 27 12:42:41 1999  
 Job time : 4 secs.



**THIS PAGE BLANK (USPTO)**



\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
\*\*\*\*\*  
(TW)

\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
\*\*\*\*\*

Mprch\_n n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Sat Nov 27 12:41:44 1999; MasPar time 26.50 Seconds  
Tabular output not generated. 153.615 Million cell updates/sec

Title: >US-09-103-287-5  
Description: (1-19) from US09103287.seq  
Perfect Score: 19  
N.A. Sequence: 1 CTTCATTATGACGATGC 19  
Comp: GAAGTAATTACTGCTACG

Scoring table: TABLE default  
Gap 10

Mmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database:

n-geneseq35  
1:part1 2:part3 3:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39 40:part40 41:part41 42:part42 43:part43  
44:part44 45:part45 46:part46 47:part47 48:part48  
49:part49 50:part50 51:part51 52:part52 53:part53  
54:part54 55:part55 56:part56 57:part57 58:part58  
59:part59 60:part60

Statistics: Mean 5.030; Variance 2.946; scale 1.708

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	19	100.0	19	60	Murc polynucleotides	9.92e-02
2	19	100.0	619	48	DNA encoding a Staphy	9.92e-02
3	19	100.0	660	60	Partial nucleotide se	9.92e-02
4	19	100.0	1351	60	UDP-N-acetylmutamate:	9.92e-02
5	17	89.5	91	9	Oligonucleotide probe	1.34e+00
6	17	89.5	91	9	Oligonucleotide probe	1.34e+00
7	17	89.5	2424	58	Staphylococcus aureus	1.34e+00
8	15	78.9	5565	43	L. lactis NS3 locus p	1.65e+01
9	14	73.7	517	43	DNA encoding a Staphy	5.52e+01

Result ID	Score	Query Match	Length	DB ID	Description	Pred. No.
10	14	73.7	1770	37	T89887	5.52e+01
11	14	73.7	3088	16	T05628	5.52e+01
12	13	73.7	7801	4	Q23500	5.52e+01
13	13	68.4	1035	24	T43303	1.79e+02
14	13	68.4	1110	24	T43316	1.79e+02
15	13	68.4	1138	9	O55627	1.79e+02
16	13	68.4	1209	58	V75054	1.79e+02
17	13	68.4	1268	60	X13494	1.79e+02
18	13	68.4	1340	10	O65353	1.79e+02
19	13	68.4	1380	60	V08865	1.79e+02
20	13	68.4	1451	58	V74988	1.79e+02
21	13	68.4	1472	58	V74866	1.79e+02
22	13	68.4	1508	60	X13044	1.79e+02
23	13	68.4	1675	7	Q43208	1.79e+02
24	13	68.4	1704	31	T71313	1.79e+02
25	13	68.4	1707	3	N50411	1.79e+02
26	13	68.4	3775	58	V74549	1.79e+02
27	13	68.4	4565	6	O36024	1.79e+02
28	13	68.4	4565	6	O36024	1.79e+02
29	13	68.4	4846	58	V74587	1.79e+02
30	13	68.4	6836	46	V31259	1.79e+02
31	13	68.4	7616	12	O70194	1.79e+02
32	13	68.4	8084	60	X13109	1.79e+02
33	13	68.4	8339	58	V74486	1.79e+02
34	13	68.4	10690	47	V32226	1.79e+02
35	13	68.4	11389	58	V43334	1.79e+02
36	13	68.4	13585	17	T11549	1.79e+02
37	13	68.4	15500	60	V08874	1.79e+02
38	13	68.4	16826	58	V74357	1.79e+02
39	13	68.4	22243	58	V74475	1.79e+02
40	13	68.4	24004	60	X13009	1.79e+02
41	13	68.4	58407	48	V21210	1.79e+02
42	13	68.4	235033	51	V57926	1.79e+02
43	13	68.4	237326	50	V57903	1.79e+02
44	13	68.4	237326	50	V57903	1.79e+02

#### ALIGNMENTS

RESULT 1  
ID V96651 standard; DNA; 19 BP.  
AC V96651;  
DT 17-MAR-1999 (first entry)  
DE Murc polynucleotides amplifying primer.  
KW Murc gene; UDP-N-acetylmutamate:L-alanine ligase; Murc polypeptide;  
KW bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
KW immunogen; drug; genetic immunisation; PCR primer; ss.  
OS Synthetic.  
NC Staphylococcus aureus.  
PN EP-889123-A2.  
PD 07-JAN-1999.  
PE 26-JUN-1998; 305064.  
PR 03-JUL-1997; US-052720.  
PA (SMIRK ) SMITHKLINE BEECHAM CORP.  
FA (SMIRK ) SMITHKLINE BEECHAM PLC.  
PI Burnham MKR, Wallis NG.  
DR WPI: 99-062655/06.  
PT New isolated Murc polypeptide from staphylococcus aureus and related  
PT nucleic acid - useful in diagnosis, treatment and prevention of  
PT bacterial infections  
PS Disclosure: Page 13; 39pp; English.  
CC The invention relates to a UDP-N-acetylmutamate:L-alanine ligase (Murc  
CC polypeptide) encoded by the S. aureus Murc gene. Host cells containing  
CC an expression system comprising the Murc gene can be used for the  
CC recombinant production of the polypeptide. Agonists or the Murc  
CC polypeptide are used to treat conditions requiring increased activity or  
CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or  
CC competitive polypeptide are useful for inhibiting the polypeptide e.g.  
CC bacterial (especially S. aureus) infections. They are also useful against  
CC Helicobacter pylori infections and related cancers, ulcers and gastritis.  
CC The antibacterial agents are useful to treat in-dwelling devices for  
CC infection prevention or generally as wound treatments to prevent adhesion  
CC of bacteria to matrix proteins. The Murc polypeptide is also useful for



CC diagnosing or prognosing a (susceptibility to) disease, for raising  
CC antibodies; to identify modulators or specific receptors; in rational  
CC drug design and as an immunogen for vaccines. The Muc gene sequences are  
CC useful in antisense/ribozyme therapeutics; to detect mutant Muc gene;  
CC for chromosomal mapping; to determine bacterial serotype; and for genetic  
CC immunisation. The present sequence represents a primer used for the PCR  
CC amplification of the Muc polynucleotides.  
CC Sequence (Muc ORF) of the Muc gene.  
SQ Sequence 19 BP; 6 A; 4 C; 3 G; 6 T;  
Query Match 100.0%; Score 19; DB 60; Length 19;  
Best Local Similarity 100.0%; Pred. No. 9.92e-02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 cttcattatgaacgatgc 19  
1 cttcattatgaacgatgc 19  
QY 1 cttcattatgaacgatgc 19  
RESULT 2  
ID V53479 standard; DNA; 619 BP.  
AC V53479;  
DT 30-OCT-1998 (first entry)  
DE DNA encoding a Staphylococcus aureus protein of unknown function.  
KW Staphylococcus aureus protein; immune response induction; eye infection;  
KW antibody production; T-cell immune response; gastrointestinal infection;  
KW respiratory infection; inhibitor; bacterial infection; cardiac infection;  
KW central nervous system; kidney infection; urinary tract infection;  
KW antimicrobial compound identification; broad spectrum antibiotic;  
KW therapy; ss.  
OS Staphylococcus aureus.  
PN FP-841394-A2.  
PD 13-MAY-1998.  
PF 24-SEP-1997; 307485.  
PR 24-SEP-1996; US-027032.  
PA (SMIR ) SMITHKLINE BEECHAM CORP.  
PI Black MT, Burnham MKR, Hodgson JE, Knowles DC,  
PI Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,  
PI Ward JM;  
DR WPI: 98-252940/23.  
DR P-PSDB: W77686.  
PT New nucleic acid sequences from Staphylococcus aureus WCHN29 -  
PT useful in vaccines and for treatment of bacterial infections of e.g.  
PT respiratory tract and central nervous system  
PS Claim 1: Page 146; 390pp; English.  
CC This sequence encodes a Staphylococcus aureus protein of unknown  
CC function, and represents a DNA sequence of the invention.  
CC The DNA sequences were isolated from Staphylococcus aureus WCHN29  
CC (NCIMB 40771). Host cells containing the DNA sequences are used to  
CC produce polypeptides or fragments. The proteins are used in the treatment  
CC of disease, for inducing an immune response by administering them, to  
CC produce antibody and/or T-cell immune response. Antagonists of the  
CC proteins are used for the inhibition of bacterial polypeptides.  
CC Conditions which may be treated include bacterial infections, especially  
CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,  
CC urinary tract, skin, bones and joints. The proteins can also be used to  
CC identify antimicrobial compounds which are broad spectrum antibiotics,  
CC especially useful in the treatment of H. pylori infection.  
SQ Sequence 619 BP; 208 A; 117 C; 79 G; 215 T;  
Query Match 100.0%; Score 19; DB 48; Length 619;  
Best Local Similarity 100.0%; Pred. No. 9.92e-02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 137 cttcattatgaacgatgc 155  
1 cttcattatgaacgatgc 19  
QY 1 cttcattatgaacgatgc 19  
RESULT 3  
ID V80065 standard; DNA; 660 BP.  
AC V80065;  
PA (SMIR ) SMITHKLINE BEECHAM CORP.

DT 17-MAR-1999 (first entry)  
DE Partial nucleotide sequence of the Muc gene.  
KW Muc gene; UDP-N-acetylmuramate:L-alanine ligase; Muc polypeptide;  
KW bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
KW immunogen; drug; genetic immunisation; ds.  
OS Staphylococcus aureus.  
PN Key Location/Qualifiers  
FT CDS 2..660  
FT /\*tag= a  
FT /product= "partial Muc polypeptide"  
PN EP-889123-A2.  
PD 07-JAN-1999.  
PF 26-JUN-1998; 305064.  
PR 03-JUL-1997; US-052720.  
PA (SMIR ) SMITHKLINE BEECHAM CORP.  
PI Black MT, Burnham MKR, Wallis NG,  
PI WPI: 99-062655/06.  
DR P-PSDB: W89139.  
PT New isolated Muc polypeptide from Staphylococcus aureus and related  
PT nucleic acid - useful in diagnosis, treatment and prevention of  
PT bacterial infections  
PS Claim 2: Pages 4-5; 39pp; English.  
CC The invention relates to a UDP-N-acetylmuramate:L-alanine ligase  
CC (Muc polypeptide) encoded by the S. aureus Muc gene. Host cells  
CC containing an expression system comprising the Muc gene can be used for  
CC the recombinant production of the polypeptide. Agonists or the Muc  
CC polypeptide are used to treat conditions requiring increased activity or  
CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or  
CC competitive polypeptide are useful for inhibiting the polypeptide e.g.  
CC bacterial (especially S. aureus) infections. They are also useful against  
CC Helicobacter pylori infections and related cancers, ulcers and gastritis.  
CC The antibacterial agents are useful to treat in-dwelling devices for  
CC infection prevention or generally as wound treatments to prevent adhesion  
CC of bacteria to matrix proteins. The Muc polypeptide is also useful for  
CC diagnosing or prognosing a (susceptibility to) disease, for raising  
CC antibodies; to identify modulators or specific receptors; in rational  
CC drug design and as an immunogen for vaccines. The Muc gene sequences are  
CC useful in antisense/ribozyme therapeutics; to detect mutant Muc gene;  
CC for chromosomal mapping; to determine bacterial serotype; and for genetic  
CC immunisation. The present sequence represents a partial nucleotide  
CC sequence (Muc ORF) of the Muc gene.  
SQ Sequence 660 BP; 233 A; 83 C; 120 G; 224 T;  
Query Match 100.0%; Score 19; DB 60; Length 660;  
Best Local Similarity 100.0%; Pred. No. 9.92e-02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 512 gcatcgttcattatgaag 530  
19 gcatcgttcattatgaag 1  
Cp 19 gcatcgttcattatgaag 1  
RESULT 4  
ID V99650 standard; DNA; 1351 BP.  
AC V99650;  
DT 17-MAR-1999 (first entry)  
DE UDP-N-acetylmuramate:L-alanine ligase (Muc polypeptide) encoding DNA.  
KW Muc gene; UDP-N-acetylmuramate:L-alanine ligase; Muc polypeptide;  
KW bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
KW immunogen; drug; genetic immunisation; ds.  
OS Staphylococcus aureus.  
PN Key Location/Qualifiers  
FT CDS 22..1335  
FT /\*tag= a  
FT /gene= "Muc"  
FT /product= "UDP-N-acetylmuramate:L-alanine ligase  
EP-889123-A2.  
PD 07-JAN-1999.  
PF 26-JUN-1998; 305064.  
PR 03-JUL-1997; US-052720.  
PA (SMIR ) SMITHKLINE BEECHAM CORP.







RESULT 10  
ID T89887 standard; DNA; 1770 BP.  
AC T89887;  
DT 27-APR-1998 (first entry)  
DE Caenorhabditis f1zled gene 1 (Cfz1) encoding a Wnt receptor  
KW Wnt receptor; Caenorhabditis f1zled gene 1; Cfz1 gene;  
KW signal transduction; cancer; cell growth; cell proliferation;  
KW mammary tumour; oncogene; therapy; ds.



```

RESULT      11
ID          105628 standard; DNA; 3088 BP.
AC          T05628;
DT          15-FEB-1996 (first entry)
DE          ADP ribosylation factor gene regulatory region.
KW          ADP ribosylation factor; promoter; ARF gene; vector; barley; ds.
OS          Hordeum vulgare.
FH          Key
FT          promoter
FT          1..3027
FT          /location=/qualifiers
FT          /*tag= a
FT          /function= ADP gene promoter region
PT          EP-681028-A1.
PD          08-NOV-1985.
PR          07-APR-1985; 105287.
PR          08-APR-1984; JP-07/1048.
PA          (SAPB ) SAPPORO BREWERIES.
PI          Hirotaka N., Ito K., Kuroda H;
DR          WPI; 95-375210/49.
PT          New plant expression vectors - conty. a regulating region derived
PT          from an ADP ribosylation factor gene of a plant
PS          Claim 3; Page 7-9; 14pp; English.
CC          The barley ADP ribosylation factor (ARF) gene was isolated by
CC          differential screening of cDNA libraries prepo. from barley roots and
CC          leaves. The regulatory (promoter) region of the gene was identified
CC          and sequenced (T05628). This region has been incorporated into
CC          expression vectors that can be used to express genes in a tissue-
CC          specific, efficient manner for use in plant breeding or the
CC          prodn. of substances using plant tissues or cells.
CC          Sequence 3088 BP; 716 A; 761 C; 939 T;
CC          Sequence 3088 BP; 716 A; 761 C; 939 T;

Query Match      73.7%; Score 14; DB 16; Length 3088;
Best Local Similarity 88.9%; Pred. NO. 5.52e+01;

```

Matches	16;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
DB	761	catagttcgttaatgaag	778						
CP	18	CATCCTTCAATGAAG	1						
RESULT	12								
ID	023500	standard; DNA; 7801 BP.							
AC	023500;								
DE	26-AUG-1992	(first entry)							
DE	Zymomonas xylose isomerase expression plasmid pXIPX.								
KW	xyIA; xyIB; pZMA; pyruvate decarboxylase; promoter; marker gene;								
KW	chloramphenicol resistance; ss.								
OS	Zymomonas mobilis.								
FH	Key	Location/Qualifiers							
FH	cds	1618..2940							
FT		/*tag= a							
FT		/product= xylose isomerase							
FT		/note= "see R2720"							
FT	cds	3013..4468							
FT		/*tag= b							
FT		/standard_name= xyIB							
FT		/note= "xylokinase"							
FT	promoter	1521..1592							
FT		/*tag= c							
FT		/label= pyruvate-decarboxylase-promoter							
FT	-35_signal	1531..1543							
FT		/*tag= d							
FT	-10_signal	1559..1568							
FT		/*tag= e							
FT	misc_feature	1..219							
FT		/*tag= f							
FT		/phenotype= chloramphenicol_resistance							
FT	misc_feature	7361..7801							
FT		/*tag= g							
FT		/phenotype= chloramphenicol_resistance							
FT	misc_feature	5414..7096							
FT		/*tag= h							
FT		/label= pZ1A							
FT		/note= "when included in a plasmid this region							
FT		allows replication in Zymomonas"							
PN	J04066090-A.								
PD	02-MAR-1992.								
PE	05-JUL-1990; 176420.								
PR	05-JUL-1990; JP-176420.								
PA	(EMBL-) SHIN ENERGY SANGYO.								
PA	(NENR-) NENRYOYO ALCOHOL KAIHAT.								
DR	WPI: 92-120685/15.								
DR	P-PSDB: R27270.								
PT	Vector for expression of heterogenes in Zymomonas sp.								
PT	used for expression of xylose isomerase for prodn. of ethanol from								
PT	xylose and/or fructose								
PS	Disclosure; Fig. 1; 14pp; Japanese.								
CC	This is the sequence of plasmid pXIPX. The plasmid can be used to								
CC	transform Zymomonas cells. The transformants can be cultured to								
CC	express the xyIA/B genes. Alcohol production from such								
CC	transformants is comparable to that from yeast.								
SO	Sequence 7801 BP; 1959 A; 1910 C; 1918 G; 2014 T;								
Query Match	73.7%;	Score 14;	DB 4;	Length 7801;					
Best Local Similarity	88.9%;	Pred. No. 5.52e+01;							
Matches 16;	Conservative								



DE Vmp7 soluble variant coding sequence.  
 KM OspA: OspB: outer surface protein A; Borrelia: variable major protein 7;  
 KM Borrelia hermslii; Vmp7: surface lipoprotein; spirochete; human; antigen;  
 KM Lyme borreliosis; relapsing fever; dermatological disorder; Lyme disease;  
 KM arthritic disorder; neurological disorder; vaccine; Borrelia lipoprotein;  
 KM ds.  
 OS Borrelia hermslii.  
 PN US5571718-A.  
 PD 05-NOV-1996.  
 PF 08-SEP-1992: 941523.  
 PR 21-DEC-1990: US-632072.  
 PA (ASUY-) ASSOC UNIVERSITIES INC.  
 PI Barbour AG, Dunn JJ;  
 DR WPI: 96-505409/50.  
 P-PSDB: W08097.  
 PT Soluble recombinant forms of Borrelia lipo:proteins - useful for  
 PT vaccine prodn. for treatment of Lyme disease  
 CC Claim 4: Column 43-46: 49pp: English.  
 CC T43301-T43303 represent coding sequences for soluble recombinant Borrelia  
 CC proteins of the invention. This sequence represents the coding sequence  
 CC for the soluble recombinant variant of the Borrelia hermslii variable  
 CC major protein 7 (Vmp7). Borrelia spirochetes are responsible for a  
 CC variety of human disorders including Lyme borreliosis, and relapsing  
 CC fevers. The spirochete is transmitted to humans and animals through the  
 CC bite of a tick, and can cause serious dermatological, arthritic, this  
 CC neurological and other pathogenic disorders in an infected host. This  
 CC sequence is used to create recombinant host cells, and the encoded Vmp7  
 CC protein can be isolated from the cytosol of one of these cells without  
 CC the use of detergent. The encoded recombinant proteins can be used as  
 CC antigens for the production of vaccines against Lyme disease. The  
 CC recombinant proteins can also be used in immunoassays and other  
 CC diagnostic screening methods to detect the presence of antibodies against  
 CC Borrelia lipoproteins in the sera of infected patients.  
 CC Sequence 1035 BP: 397 A; 141 C; 274 G; 223 T;  
 SQ

Query Match 68.4%; Score 13; DB 24; Length 1035;  
 Best Local Similarity 93.3%; Pred. No. 1.79e+02;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 814 cgtcattagtgaag 828  
 |||||||  
 CP 15 CGTCATTATGAG 1

RESULT 14  
 ID T43316 standard; DNA: 1110 BP.  
 AC T43316;  
 DT 10-FEB-1997 (first entry)  
 DE Vmp7 variant #1 coding sequence.  
 KM OspA: OspB: outer surface protein A; Borrelia: variable major protein 7;  
 KM Borrelia hermslii; Vmp7: surface lipoprotein; spirochete; human; antigen;  
 KM Lyme borreliosis; relapsing fever; dermatological disorder; Lyme disease;  
 KM arthritic disorder; neurological disorder; vaccine; Borrelia lipoprotein;  
 KM ds.  
 OS Synthetic.  
 PN US5571718-A.  
 PD 05-NOV-1996.  
 PF 08-SEP-1992: 941523.  
 PR 21-DEC-1990: US-632072.  
 PA (ASUY-) ASSOC UNIVERSITIES INC.  
 PI Barbour AG, Dunn JJ;  
 DR WPI: 96-505409/50.  
 PT Soluble recombinant forms of Borrelia lipo:proteins - useful for  
 PT vaccine prodn. for treatment of Lyme disease  
 CC Example 9: Column 41-42: 49pp: English.  
 CC This sequence represents the coding sequence for a soluble recombinant  
 CC Borrelia variable major protein 7 (Osp7) variant protein of the  
 CC invention. Borrelia spirochetes are responsible for a variety of human  
 CC disorders including Lyme borreliosis, and relapsing fevers. The  
 CC spirochete is transmitted to humans and animals through the bite of a  
 CC tick, and can cause serious dermatological, arthritic, neurological and

CC other pathogenic disorders in an infected host. This sequence is used to  
 CC create recombinant host cells, and the encoded Vmp7 protein can be  
 CC isolated from the cytosol of one of these cells without the use of  
 CC detergent. The encoded recombinant proteins can be used as antigens for  
 CC the production of vaccines against Lyme disease. The recombinant  
 CC proteins can also be used in immunoassays and other diagnostic screening  
 CC methods to detect the presence of antibodies against Borrelia  
 CC lipoproteins in the sera of infected patients.  
 CC Sequence 1110 BP: 432 A; 143 C; 286 G; 249 T;  
 SQ

Query Match 68.4%; Score 13; DB 24; Length 1110;  
 Best Local Similarity 100.0%; Pred. No. 1.79e+02;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 889 cgtcattagtgaag 903  
 |||||||  
 CP 15 CGTCATTATGAG 1

RESULT 15  
 ID Q55627 standard; DNA: 1138 BP.  
 AC Q55627;  
 DT 12-JUL-1994 (first entry)  
 DE Human GM-CSF gene.  
 KM Granulocyte Macrophage Colony Stimulating Factor; GM-CSF;  
 KM non-coding region; coding region; resonance; interaction;  
 KM optimization; promoter region; TATA box; ss.  
 OS Homo sapiens.  
 PN FR2692594-A.  
 PD 24-DEC-1993.  
 PR 22-JUN-1992: 007571.  
 PA (PERE/) PEREZ J.  
 PI Perez J;  
 DR WPI: 94-028256/04.  
 PT Application of optimised gene expression - for scientific,  
 PT industrial and therapeutic purposes  
 PS Disclosure: Fig 13; 110pp; French.  
 CC Alterations were made to the TATA box sequence of the human and  
 CC mouse GM-CSF genes (see Q55627 and Q55628, respectively) and the  
 CC effects of the various alterations on gene expression were measured.  
 CC Results indicated that resonances of the LRF type are most strongly  
 CC affected by changes in the TATA box; it was further shown that  
 CC LRF resonances control expression.  
 CC Sequence 1138 BP: 274 A; 306 C; 292 G; 266 T;  
 SQ

Query Match 68.4%; Score 13; DB 9; Length 1138;  
 Best Local Similarity 100.0%; Pred. No. 1.79e+02;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 177 gtccattaatgaa 189  
 |||||||  
 CP 14 GTTCATTATGAA 2

Search completed: Sat Nov 27 12:42:19 1999  
 Job time : 35 secs.



\*\*\*\*\*  
WIDEORL (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*  
Merch\_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 12:42:58 1999; Maspar time 6.06 Seconds  
271.193 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-5  
Description: (1-19) from US09103287.seq  
Perfect Score: 19  
N.A. Sequence: 1 CTTGATATGACGATGC 19  
Comp: GAAGTATATGCTGCTACG

Scoring table: TABLE default  
Gap 10

Nmatch STD : Dbase 0; Query 0

Searched: 165359 seqs, 43243793 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: n-issued  
1:5A\_COMB 2:5B\_COMB 3:5C\_COMB 4:PCT9\_COMB 5:backfiles1

Statistics: Mean 4.735; Variance 2.435; scale 1.945

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
C 1	14	73.7	3088	3	US-08-418-	Sequence 1, Applicatio	6.63e+00
C 2	13	68.4	913	1	US-08-109-	Sequence 1, Applicatio	2.60e+01
C 3	13	68.4	913	2	US-08-459-	Sequence 1, Applicatio	2.60e+01
C 4	13	68.4	1035	1	US-07-941-	Sequence 22, Applicatio	2.60e+01
C 5	13	68.4	1110	1	US-07-941-	Sequence 20, Applicatio	2.60e+01
C 6	13	68.4	1340	1	US-07-971-	Sequence 1, Applicatio	2.60e+01
C 7	13	68.4	1675	4	PCT-US92-1	Sequence 12, Applicatio	2.60e+01
C 8	13	68.4	4080	2	PCT-US95-1	Sequence 1, Applicatio	2.60e+01
C 9	13	68.4	4080	2	US-08-375-	Sequence 1, Applicatio	2.60e+01
C 10	13	68.4	4129	3	US-08-370-	Sequence 12, Applicatio	2.60e+01
C 11	13	68.4	7616	4	PCT-US94-0	Sequence 54, Applicatio	2.60e+01
C 12	12	63.2	47	1	US-08-441-	Sequence 60, Applicatio	9.75e+01
C 13	12	63.2	287	3	US-08-418-	Sequence 1, Applicatio	9.75e+01
C 14	12	63.2	485	2	US-08-385-	Sequence 1, Applicatio	9.75e+01
C 15	12	63.2	485	1	US-08-240-	Sequence 13, Applicatio	9.75e+01
C 16	12	63.2	730	1	US-08-332-	Sequence 49, Applicatio	9.75e+01
C 17	12	63.2	832	1	US-08-374-	Sequence 11, Applicatio	9.75e+01
C 18	12	63.2	872	1	US-08-374-	Sequence 10, Applicatio	9.75e+01
C 19	12	63.2	1028	1	US-08-029-	Sequence 1, Applicatio	9.75e+01
C 20	12	63.2	1200	2	US-08-672-	Sequence 2, Applicatio	9.75e+01

C 21	12	63.2	1358	3	US-08-454-	Sequence 20, Applicatio	9.75e+01
C 22	12	63.2	1358	1	US-08-444-	Sequence 20, Applicatio	9.75e+01
C 23	12	63.2	1358	3	US-08-455-	Sequence 20, Applicatio	9.75e+01
C 24	12	63.2	1358	3	US-08-456-	Sequence 20, Applicatio	9.75e+01
C 25	12	63.2	1521	2	US-08-753-	Sequence 1, Applicatio	9.75e+01
C 26	12	63.2	1521	2	US-08-753-	Sequence 1, Applicatio	9.75e+01
C 27	12	63.2	1989	2	US-08-454-	Sequence 5, Applicatio	9.75e+01
C 28	12	63.2	2229	5	US-08-453-	Sequence 5, Applicatio	9.75e+01
C 29	12	63.2	2490	1	Patent No. 5470718-	Sequence 45, Applicatio	9.75e+01
C 30	12	63.2	2694	3	US-08-332-	Sequence 38, Applicatio	9.75e+01
C 31	12	63.2	3466	3	US-08-376-	Sequence 3, Applicatio	9.75e+01
C 32	12	63.2	3738	1	US-08-304-	Sequence 7, Applicatio	9.75e+01
C 33	12	63.2	4483	2	US-08-449-	Sequence 5, Applicatio	9.75e+01
C 34	12	63.2	4483	3	US-08-454-	Sequence 5, Applicatio	9.75e+01
C 35	12	63.2	4483	3	US-08-444-	Sequence 5, Applicatio	9.75e+01
C 36	12	63.2	4483	3	US-08-455-	Sequence 5, Applicatio	9.75e+01
C 37	12	63.2	4483	1	US-08-181-	Sequence 5, Applicatio	9.75e+01
C 38	12	63.2	4507	3	US-08-568-	Sequence 3, Applicatio	9.75e+01
C 39	12	63.2	4615	3	US-08-574-	Sequence 3, Applicatio	9.75e+01
C 40	12	63.2	4699	2	US-08-449-	Sequence 6, Applicatio	9.75e+01
C 41	12	63.2	4699	1	US-08-449-	Sequence 6, Applicatio	9.75e+01
C 42	12	63.2	4699	1	US-08-444-	Sequence 6, Applicatio	9.75e+01
C 43	12	63.2	6854	2	US-08-468-	Sequence 4, Applicatio	9.75e+01
C 44	12	63.2	7101	2	US-08-480-	Sequence 9, Applicatio	9.75e+01
C 45	12	63.2	8775	2	US-08-232-	Sequence 5, Applicatio	9.75e+01

## ALIGNMENTS

RESULT 1  
ID US-08-418-444A-1 STANDARD; DNA; UNC; 3088 BP.  
AC xxxxxx  
DT  
DE Sequence 1, Application US/08418444A  
CC Sequence 1, Application US/08418444A  
CC Patent No. 5773688  
CC  
CC GENERAL INFORMATION:  
CC APPLICANT: KURODA, HISAO  
CC APPLICANT: HIROTA, NAOHICO  
CC APPLICANT: ITO, KAZUTOSHI  
CC TITLE OF INVENTION: GENE EXPRESSION VECTOR USING THE GENE  
CC TITLE OF INVENTION: EXPRESSION REGULATING REGION OF THE ADP RIBOSYLATION  
CC NUMBER OF SEQUENCES: 9  
CC  
CC CORRESPONDENCE ADDRESSES:  
CC ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT  
CC STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CC CITY: ARLINGTON  
CC STATE: VIRGINIA  
CC COUNTRY: USA  
CC ZIP: 22202  
CC  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.30  
CC  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/418,444A  
CC FILING DATE: 07-APR-1995  
CC  
CC CLASSIFICATION: 800  
CC  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: JP HEI 6-71048  
CC FILING DATE: 08-APR-1994  
CC  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: OBLON, NO. 5773688man F.  
CC REGISTRATION NUMBER: 24,618  
CC REFERENCE/DOCKET NUMBER: 2589-024-0  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (703) 413-3000  
CC TELEFAX: (703) 413-2220  
CC  
CC TELEX: 248855 OPAT UR  
CC  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:



CC LENGTH: 3088 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
SQ SEQUENCE 3088 BP; 716 A; 761 C; 672 G; 939 T; 0 OTHER.

Query Match 73.7%; Score 14; DB 3; Length 3088;  
Best Local Similarity 88.9%; Pred. No. 6.63e+00;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 761 CATGTCGTTAATGAG 778  
Cp 18 CATGTCGTTAATGAG 1

RESULT 2  
ID US-08-109-391A-1 STANDARD; DNA; UNC; 913 BP.  
AC xxxxxx

DE Sequence 1, Application US/08109391A  
CC Sequence 1, Application US/08109391A  
CC Patent No. 5639876  
CC GENERAL INFORMATION:  
CC APPLICANT: Tripp, Cynthia A.  
CC APPLICANT: Frank, Glenn R.  
CC TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING NOVEL  
CC TITLE OF INVENTION: PARASITIC HELMINTH PROTEINS  
CC NUMBER OF SEQUENCES: 17  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Sheridan Ross & McIntosh  
CC STREET: 1700 Lincoln St., Suite 3500  
CC CITY: Denver  
CC STATE: CO  
CC COUNTRY: U.S.A.  
CC ZIP: 80203  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/109,391A  
CC FILING DATE: 19-AUG-1993  
CC CLASSIFICATION: 536  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Connell, Gary J.  
CC REGISTRATION NUMBER: 32,020  
CC REFERENCE/DOCKET NUMBER: 2618-13  
CC TELEPHONE: 303/863-9700  
CC TELEFAX: 303/863-0223  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 913 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: CDNA  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 3..911  
SQ SEQUENCE 913 BP; 363 A; 115 C; 223 G; 212 T; 0 OTHER.

Query Match 68.4%; Score 13; DB 1; Length 913;  
Best Local Similarity 93.3%; Pred. No. 2.60e+01;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 250 CATGTCGTTAATG 264  
Cp 18 CATGTCGTTAATG 4

RESULT 3  
ID US-08-459-019A-1 STANDARD; DNA; UNC; 913 BP.  
AC xxxxxx

DE Sequence 1, Application US/08459019A  
CC Sequence 1, Application US/08459019A  
CC Patent No. 5686080  
CC GENERAL INFORMATION:  
CC APPLICANT: Tripp, Cynthia A.  
CC APPLICANT: Frank, Glenn R.  
CC APPLICANT: Griev, Robert B.  
CC TITLE OF INVENTION: NOVEL PARASITIC HELMINTH P4 PROTEINS  
CC NUMBER OF SEQUENCES: 17  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Sheridan Ross & McIntosh  
CC STREET: 1700 Lincoln Street, #3500  
CC CITY: Denver  
CC STATE: CO  
CC COUNTRY: U.S.A.  
CC ZIP: 80203  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/459,019A  
CC FILING DATE: 06-JUN-1995  
CC CLASSIFICATION: 536  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Connell, Gary J.  
CC REGISTRATION NUMBER: 33,020  
CC REFERENCE/DOCKET NUMBER: 2618-13-1  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (303) 863-9700  
CC TELEFAX: (303) 863-0223  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 913 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: CDNA  
CC HYPOTHETICAL: NO  
CC ANTI-SENSE: NO  
CC ORIGINAL SOURCE:  
CC ORGANISM: Dirofilaria immitis  
CC DEVELOPMENTAL STAGE: Larva  
CC IMMEDIATE SOURCE:  
CC LIBRARY: L3 and/or L4 larval D. immitis CDNA  
CC LIBRARY: expression library  
CC CLONE: P4  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 3..911  
SQ SEQUENCE 913 BP; 363 A; 115 C; 223 G; 212 T; 0 OTHER.

Query Match 68.4%; Score 13; DB 2; Length 913;  
Best Local Similarity 93.3%; Pred. No. 2.60e+01;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 250 CATGTCGTTAATG 264  
Cp 18 CATGTCGTTAATG 4

RESULT 4  
ID US-07-941-523-22 STANDARD; DNA; UNC; 1035 BP.  
AC xxxxxx

DE Sequence 22, Application US/07941523  
CC Sequence 22, Application US/07941523



CC Patent No. 5571718  
CC GENERAL INFORMATION:  
CC APPLICANT: Dunn, John J  
CC APPLICANT: Barbour, Alan G  
CC TITLE OF INVENTION: Cloning and Expression of Borrelia  
CC TITLE OF INVENTION: Lipoproteins  
CC NUMBER OF SEQUENCES: 24  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
CC STREET: Two Millitia Drive  
CC CITY: Lexington  
CC STATE: Massachusetts  
CC COUNTRY: U.S.A.  
CC ZIP: 01730  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/941,523  
CC FILING DATE: 19920908  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Granahan, Patricia  
CC REGISTRATION NUMBER: 32,227  
CC REFERENCE/DOCKET NUMBER: BNL90-01A  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 861-6240  
CC TELEFAX: (617) 861-9540  
CC INFORMATION FOR SEQ ID NO: 22:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1035 base pairs  
CC TYPE: NUCLEIC ACID  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC SEQ SEQUENCE 1035 BP; 397 A; 141 C; 274 G; 223 T; 0 OTHER.

Query Match 68.4%; Score 13; DB 1; Length 1035;  
Best Local Similarity 93.3%; Pred. No. 2.60e+01;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 814 CGTTCATTAGTGAAG 828  
|||||  
Cp 15 CGTTCATTATGAAG 1

RESULT 5  
ID US-07-941-523-20 STANDARD; DNA; UNC; 1110 BP.  
AC xxxxxx  
DT Sequence 20, Application US/07941523  
DE Sequence 20, Application US/07941523  
CC Patent No. 5571718  
CC GENERAL INFORMATION:  
CC APPLICANT: Dunn, John J  
CC APPLICANT: Barbour, Alan G  
CC TITLE OF INVENTION: Cloning and Expression of Borrelia  
CC TITLE OF INVENTION: Lipoproteins  
CC NUMBER OF SEQUENCES: 24  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
CC STREET: Two Millitia Drive  
CC CITY: Lexington  
CC STATE: Massachusetts  
CC COUNTRY: U.S.A.  
CC ZIP: 01730  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/941,523  
CC FILING DATE: 19920908  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Granahan, Patricia  
CC REGISTRATION NUMBER: 32,227  
CC REFERENCE/DOCKET NUMBER: BNL90-01A  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 861-6240  
CC TELEFAX: (617) 861-9540  
CC INFORMATION FOR SEQ ID NO: 20:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1110 base pairs  
CC TYPE: NUCLEIC ACID  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC SEQ SEQUENCE 1110 BP; 432 A; 143 C; 286 G; 249 T; 0 OTHER.

Query Match 68.4%; Score 13; DB 1; Length 1110;  
Best Local Similarity 93.3%; Pred. No. 2.60e+01;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 889 CGTTCATTAGTGAAG 903  
|||||  
Cp 15 CGTTCATTATGAAG 1

RESULT 6  
ID US-07-971-558-1 STANDARD; DNA; UNC; 1340 BP.  
AC xxxxxx  
DT Sequence 1, Application US/07971558  
DE Sequence 1, Application US/07971558  
CC Patent No. 5441736  
CC GENERAL INFORMATION:  
CC APPLICANT: GERLACH, GERALD F.  
CC APPLICANT: WILLSON, PHILIP J.  
CC APPLICANT: ROSSI-CAMPOS, AMALIA  
CC APPLICANT: PORTER, ANDREW A.  
CC TITLE OF INVENTION: ACTINORACILLUS PLEUROSPONGINAE OUTER  
CC TITLE OF INVENTION: MEMBRANE LIPOPROTEIN A AND USBS THEREOF  
CC NUMBER OF SEQUENCES: 2  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: ROBERTA L. ROBINS  
CC STREET: 635 BRYANT STREET  
CC CITY: PALO ALTO  
CC STATE: CALIFORNIA  
CC COUNTRY: UNITED STATES OF AMERICA  
CC ZIP: 94301  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/971,558  
CC FILING DATE: 19921105  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: ROBINS, ROBERTA L.  
CC REGISTRATION NUMBER: 33,208  
CC REFERENCE/DOCKET NUMBER: 9000-0027  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 617-8999  
CC TELEFAX: (415) 327-3231  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1340 base pairs  
CC TYPE: NUCLEIC ACID  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear



CC MOLECULE TYPE: DNA (genomic)

CC FEATURE: NAME/KEY: CDS

CC LOCATION: 158..1252

CC SEQUENCE 1340 BP: 502 A; 183 C; 265 G; 390 T; 0 OTHER.

SQ QUERY MATCH 68.4%; Score 13; DB 1; Length 1340;

Best Local Similarity 93.3%; Pred. No. 2.60e+01;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 823 ATCAATGACGATGC 837

Gy 5 ATTATGACGATGC 19

RESULT 7

ID PCT-US92-10284-12 STANDARD; DNA: UNC; 1675 BP.

AC xxxxxx

DE Sequence 12, Application PC/TUS9210284

CC Sequence 12, Application PC/TUS9210284

CC GENERAL INFORMATION:

CC APPLICANT: Browne, John, Kinney, Anthony J.,

CC APPLICANT: Pierce, John, Wierzbicki, Anna M.,

CC APPLICANT: Vaday, Narendra S., Perez-Grau, Luis

CC TITLE OF INVENTION: Fatty Acid Desaturase Genes

CC TITLE OF INVENTION: from Plants

CC NUMBER OF SEQUENCES: 32

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: E. I. du Pont de Nemours and Company

CC STREET: 1007 Market Street

CC CITY: Wilmington

CC STATE: Delaware

CC COUNTRY: U.S.A.

CC ZIP: 19898

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC OPERATING SYSTEM: Macintosh System, 6.0

CC SOFTWARE: Microsoft Word, 4.0

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: PCT/US92/10284

CC FILING DATE: 19921203

CC CLASSIFICATION: 435

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 07/804,259

CC FILING DATE: 4 DECEMBER 1991

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Floyd, Linda A.

CC REGISTRATION NUMBER: 33,692

CC REFERENCE/DOCKET NUMBER: BB-1036-A

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (302) 992-4929

CC TELEFAX: (302) 892-7949

CC TELEX: 835420

CC INFORMATION FOR SEQ ID NO: 12:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 1675 base pairs

CC TYPE: NUCLEIC ACID

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: CDNA

CC HYPOTHETICAL: NO

CC ORIGINAL SOURCE:

CC ORGANISM: Glycine max

CC IMMEDIATE SOURCE:

CC CLONE: pSPD-118bwp

CC FEATURE:

CC NAME/KEY: CDS

CC LOCATION: 169..1530

CC SEQUENCE 1675 BP: 417 A; 329 C; 375 G; 554 T; 0 OTHER.

SQ QUERY MATCH 68.4%; Score 13; DB 4; Length 1675;

Best Local Similarity 88.2%; Pred. No. 2.60e+01;

Db 1639 ATCTTCATTATAAG 1655

Cp 17 ATCTTCATTATAAG 1

RESULT 8

ID PCT-US95-16930-1 STANDARD; DNA: UNC; 4080 BP.

AC xxxxxx

DE Sequence 1, Application PC/TUS9516930

CC Sequence 1, Application PC/TUS9516930

CC GENERAL INFORMATION:

CC APPLICANT: UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL

CC TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE

CC TITLE OF INVENTION: PRODUCTION IN THE ABSENCE OF

CC TITLE OF INVENTION: NONSENSE-MEDIATED MRNA DECAY

CC TITLE OF INVENTION: FUNCTION

CC NUMBER OF SEQUENCES: 6

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Fish &amp; Richardson

CC STREET: 225 Franklin Street Suite 3100

CC CITY: Boston

CC STATE: MA

CC COUNTRY: USA

CC ZIP: 02110-2804

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC OPERATING SYSTEM: IBM PC compatible

CC SOFTWARE: Patent Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: PCT/US95/16930

CC FILING DATE: 27-DEC-1995

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/375,300

CC FILING DATE: 20-JAN-1995

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Fasse, J. Peter

CC REGISTRATION NUMBER: 32,983

CC REFERENCE/DOCKET NUMBER: 04020/046W01

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617)542-5070

CC TELEFAX: (617)542-8906

CC INFORMATION FOR SEQ ID NO: 1:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 4080 base pairs

CC TYPE: nucleic acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: DNA

SQ SEQUENCE 4080 BP: 1449 A; 645 C; 801 G; 1185 T; 0 OTHER.

Query Match 68.4%; Score 13; DB 4; Length 4080;

Best Local Similarity 88.2%; Pred. No. 2.60e+01;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 635 CTCTATTATGAAGAT 651

Gy 1 CTCTATTATGAAGAT 17

RESULT 9

ID US-08-375-300-1 STANDARD; DNA: UNC; 4080 BP.

AC xxxxxx

DE Sequence 1, Application US/08375300

CC Sequence 1, Application US/08375300

CC Patent No. 5679566

CC GENERAL INFORMATION:



CC APPLICANT: Feng, He  
CC APPLICANT: Jacobson, Allan S.  
CC TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN  
CC TITLE OF INVENTION: THE ABSENCE OF NONSENSE-MEDIATED MRNA DECAY FUNCTION  
CC NUMBER OF SEQUENCES: 6  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Fish & Richardson  
CC STREET: 225 Franklin Street suite 3100  
CC CITY: Boston  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 02110-2804  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: IBM PC compatible  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/375,300  
CC FILING DATE: 20-JAN-1995  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Fasse, J. P.  
CC REGISTRATION NUMBER: 32,983  
CC REFERENCE/DOCKET NUMBER: 04020/046001  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617)542-5070  
CC TELEFAX: (617)542-8906  
CC TELEEX: 200154  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 4080 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA  
SQ SEQUENCE 4080 BP; 1449 A; 645 C; 801 G; 1185 T; 0 OTHER.

Query Match 68.4%; Score 13; DB 2; Length 4080;  
Best Local Similarity 88.2%; Pred. No. 2.60e+01;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 635 CTCTCTTTATGAGAT 651  
|||||  
QY 1 CTTCATTATGACGAT 17

RESULT 10  
ID US-08-370-319C-12 STANDARD: DNA; UNC; 4129 BP.

DE Sequence 12, Application US/08370319C  
CC Sequence 12, Application US/08370319C  
CC Patent No. 5856091  
CC GENERAL INFORMATION:  
CC APPLICANT: Brichard, Vincent; Van Pel, Aline;  
CC APPLICANT: Traversari, Catia; W lfel, Thomas; Coullie, Pierre;  
CC APPLICANT: Boon-Falleur, Thierry; De Plaen, Etienne  
CC TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A  
CC TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR PROCESSED TO AT LEAST ONE  
CC TITLE OF INVENTION: REJECTION ANTIGEN PRESENTED BY HLA-A2  
CC NUMBER OF SEQUENCES: 12  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Felfe & Lynch  
CC STREET: 805 Third Avenue  
CC CITY: New York City  
CC STATE: New York  
CC COUNTRY: USA  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
CC OPERATING SYSTEM: PC-DOS  
CC SOFTWARE: Wordperfect

CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/370,319C  
CC FILING DATE: 10-JANUARY-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/272,351  
CC FILING DATE: 8-JULY-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/032,978  
CC FILING DATE: 18-MAR-1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Hanson, No. 5856091man D.  
CC REGISTRATION NUMBER: 30,946  
CC REFERENCE/DOCKET NUMBER: LUD 5377.1  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 688-9200  
CC TELEFAX: (212) 838-3884  
CC INFORMATION FOR SEQ ID NO: 12:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 4129 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC FEATURE:  
CC OTHER INFORMATION: The sequence is preceded by an  
CC OTHER INFORMATION: unsequenced portion of from 4.7 to 5.3  
CC OTHER INFORMATION: kilobases  
SQ SEQUENCE 4129 BP; 1300 A; 827 C; 728 G; 1273 T; 1 OTHER.

Query Match 68.4%; Score 13; DB 3; Length 4129;  
Best Local Similarity 93.3%; Pred. No. 2.60e+01;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 2710 ATGGTCATTATGA 2724  
|||  
Cp 17 ATCGTCATTATGA 3

RESULT 11  
ID PCT-US94-01149-54 STANDARD: DNA; UNC; 7616 BP.

DE Sequence 54, Application PC/TUS9401149  
CC Sequence 54, Application PC/TUS9401149  
CC GENERAL INFORMATION:  
CC APPLICANT: Shatzman, Allan  
CC APPLICANT: Scott, Miller  
CC APPLICANT: Dillon, Susan B.  
CC APPLICANT: Kane, James  
CC TITLE OF INVENTION: Vaccinal Polypeptides  
CC NUMBER OF SEQUENCES: 72  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: SmithKline Beecham Corporation - Corporate  
CC STREET: U.S. Mailcode UW2220 - 709 Swedeland Road  
CC CITY: King of Prussia  
CC STATE: Pennsylvania  
CC COUNTRY: USA  
CC ZIP: 19406-2799  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US94/01149  
CC FILING DATE:  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 149,150  
CC FILING DATE: 05-NOV-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 013,415







CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/418,071  
CC FILING DATE: 06-APR-1995  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Haley Jr., James F.  
CC REGISTRATION NUMBER: 27,794  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 596-9000  
CC TELEFAX: (212) 596-9090  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 287 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: subgenomic DNA  
CC DESCRIPTION: /desc="clone 1 insert"  
CC HYPOTHETICAL: NO  
CC ANTI-SENSE: NO  
CC ORIGINAL SOURCE:  
CC ORGANISM: Banana Bunchy Top Virus (BBTV)  
SQ SEQUENCE 287 BP; 98 A; 48 C; 76 G; 65 T; 0 OTHER.

Query Match 63.2%; Score 12; DB 3; Length 287;  
Best Local Similarity 100.0%; Pred. No. 9.75e+01;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 236 TCATTATGAG 247  
Cp 12 TCATTATGAG 1

RESULT 14  
ID US-08-385-187A-1 STANDARD; DNA; UNC; 485 BP.  
AC xxxxxx  
DT Sequence 1, Application US/08385187A  
DE Sequence 1, Application US/08385187A  
CC Patent No. 5723310  
CC GENERAL INFORMATION:  
CC APPLICANT: Builder, Stuart  
CC APPLICANT: Hart, Roger  
CC APPLICANT: Lester, Philip  
CC APPLICANT: Ogez, John  
CC APPLICANT: Reifsnnyder, David  
CC TITLE OF INVENTION: Aqueous Multiple-Phase Isolation of  
CC NUMBER OF SEQUENCES: 9  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Genentech, Inc.  
CC STREET: 460 Point San Bruno Blvd  
CC CITY: South San Francisco  
CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 94080  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Winpatin (Genentech)  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/385,187A  
CC FILING DATE: 07-Feb-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US94/09089  
CC FILING DATE: 10-AUG-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/110663  
CC FILING DATE: 20-AUG-1993  
CC ATTORNEY/AGENT INFORMATION:

CC NAME: Hasak, Janet E.  
CC REGISTRATION NUMBER: 28,616  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 415/225-1896  
CC TELEFAX: 415/952-9881  
CC TELEX: 910/371-7168  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 485 base pairs  
CC TYPE: Nucleic Acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
SQ SEQUENCE 485 BP; 117 A; 113 C; 114 G; 141 T; 0 OTHER.

Query Match 63.2%; Score 12; DB 2; Length 485;  
Best Local Similarity 87.5%; Pred. No. 9.75e+01;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 386 GCATCGTTCATGAG 401  
Cp 19 GCATCGTTCATGAG 4

RESULT 15  
ID US-08-240-121-13 STANDARD; DNA; UNC; 485 BP.  
AC xxxxxx  
DT Sequence 13, Application US/08240121  
DE Sequence 13, Application US/08240121  
CC Patent No. 5487980  
CC GENERAL INFORMATION:  
CC APPLICANT: Swartz, James  
CC TITLE OF INVENTION: Method for Producing Polypeptide via Bacterial Fermentat  
CC NUMBER OF SEQUENCES: 21  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Genentech, Inc.  
CC STREET: 460 Point San Bruno Blvd  
CC CITY: South San Francisco  
CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 94080  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: patin (Genentech)  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/240,121  
CC FILING DATE: 09-May-1994  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/989844  
CC FILING DATE: 23-NOV-5487980-1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Hasak, Janet E.  
CC REGISTRATION NUMBER: 28,616  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 415/225-1896  
CC TELEFAX: 415/952-9881  
CC TELEX: 910/371-7168  
CC INFORMATION FOR SEQ ID NO: 13:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 485 bases  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
SQ SEQUENCE 485 BP; 117 A; 113 C; 114 G; 141 T; 0 OTHER.

Query Match 63.2%; Score 12; DB 1; Length 485;  
Best Local Similarity 87.5%; Pred. No. 9.75e+01;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;



Mon Nov 29 13:18:21 1999

US-09-103--287-5.rni

Page 8

Db 386 GCATCGTTCATGATG 401  
|||||||  
Cp 19 GCATCGTTCATGATG 4

Search completed: Sat Nov 27 12:43:09 1999  
Job time : 11 secs.







OT F I N E R C

RESULT 2

ENTRY S16901 #type complete

TITLE cytochrome P450 19 - rat

ALTERNATE\_NAMES cytochrome P450arom

ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat

DATE 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 05-Mar-1999

ACCESSIONS S16901

REFERENCE S16901

#authors Lepoint, E.D.; Peterson, K.G.; Noble, J.F.; George, F.W.; McPhaul, M.J.

#journal Mol. Cell. Endocrinol. (1990) 70:31-40

#title The structure of cDNA clones encoding the aromatase P-450 isolated from a rat Leydig cell tumor line demonstrates differential processing of aromatase mRNA in rat ovary and a neoplastic cell line.

#cross-references M01D:90255798

#accession S16901

#status preliminary

#molecule\_type mRNA

GENETICS ##residues 1-496 ##label LEP

CLASSIFICATION CYP19

KEYWORDS #superfamily human cytochrome P450 CYP19; cytochrome P450 homology

FEATURE oxidoreductase; heme; iron; microsome; monooxygenase; oxidoreductase; transmembrane protein

436 #binding-site heme iron (Cys) (axial ligand) #status predicted

SUMMARY #length 496 #molecular-weight 56663 #checksum 3858

Query Match 80.0%; Score 76; DB 2; Length 496;

Best Local Similarity 52.6%; Pred. No. 5.60e+00;

Matches 10; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

GENETICS

#gene

CLASSIFICATION

ENTRY F I N E S C

TITLE 1071 YTTVATHAAYGARMSNTGY 1089

ORGANISM :||:||||:||||: ||:

DATE 1 CTTCAATTATGAACGATGC 19

OT F I N E R C

RESULT 3

ENTRY D64044 #type complete

TITLE hypothetical protein HI0036 - Haemophilus influenzae (strain Rd KW20)

ORGANISM #formal\_name Haemophilus influenzae

DATE 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 18-Sep-1999

ACCESSIONS D64044

REFERENCE A64000

#authors Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.; Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Sauder, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.

#journal Science (1995) 269:496-512

#title Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

#cross-references M01D:95350630

#accession D64044

#status nucleic acid sequence not shown; translation not shown

#molecule\_type DNA

##residues 1-592 ##label TIGR

##cross-references GB:U32689; GB:I42023; NID:g1572980; PID:g1572982; TIGR:HI0036

CLASSIFICATION #superfamily ATP-binding cassette homology

FEATURE

ENTRY 407-585

TITLE #domain ATP-binding cassette homology #label ABC\

ALTERNATE\_NAMES 424-431

ORGANISM #region nucleotide-binding motif A (P-loop)

SUMMARY #length 592 #molecular-weight 67903 #checksum 1875

Query Match 80.0%; Score 76; DB 2; Length 592;

Best Local Similarity 52.9%; Pred. No. 5.60e+00;

Matches 9; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

DB F I N E R

ENTRY 723 YTTVATHAAYGARMCNY 739

TITLE :||:||||:||||: ||:

ORGANISM 1 CTTCAATTATGAACGAT 17

DATE F I N E R

RESULT 4

ENTRY S65799 #type complete

TITLE chromosome scaffold protein suda - Emeritella nidulans

ALTERNATE\_NAMES #formal\_name Emeritella nidulans, Aspergillus nidulans

DATE 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 13-Sep-1998

ACCESSIONS S65799

REFERENCE S65799

#authors Holt, C.L.; May, G.S.

#journal Genetics (1996) 142:777-787

#title An extragenic suppressor of the mitosis-defective bim6 mutation of Aspergillus nidulans codes for a chromosome scaffold protein.

#accession S65799

#status preliminary; nucleic acid sequence not shown

#molecule\_type mRNA

GENETICS ##residues 1-1211 ##label HOL

##cross-references EMBL:U40146; NID:g1103892; PID:g1103893

CLASSIFICATION

ENTRY suda

TITLE #superfamily hypothetical protein YJL074C

ORGANISM #length 1211 #molecular-weight 138958 #checksum 5614

SUMMARY

Query Match 80.0%; Score 76; DB 2; Length 1211;

Best Local Similarity 52.9%; Pred. No. 5.60e+00;

Matches 9; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

GENETICS

#gene

CLASSIFICATION

ENTRY F I N E R

TITLE 582 YTTVATHAAYGARMCNY 598

ORGANISM :||:||||:||||: ||:

DATE 1 CTTCAATTATGAACGAT 17

OT F I N E R

RESULT 5

ENTRY F64456 #type complete

TITLE hypothetical protein MJ1255 - Methanococcus jannaschii

ALTERNATE\_NAMES #formal\_name Methanococcus jannaschii

DATE 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 10-Oct-1997

ACCESSIONS F64456

REFERENCE A64300

#authors Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghagen, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

#journal Science (1996) 273:1058-1073

#title Complete genome sequence of the methanogenic archaeon,



Methanococcus jannaschii.

#cross-references MUID:96337999

#accession F64456

#status

preliminary: nucleic acid sequence not shown;  
translation not shown

##molecule\_type DNA

##residues 1-398 #label BUL

##cross-references GB:U67566; GB:U77117; NID:g1591887; PID:g1591889;  
TIGR:MJ1255; PID:g1511260

## GENETICS

#map\_position FOR197420-1198616

#start\_codon TGC

SUMMARY #length 398 #molecular-weight 46068 #checksum 4354

Query Match 75.8%; Score 72; DB 2; Length 398;

Best Local Similarity 58.8%; Pred. No. 2.53e+01;  
Matches 10; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Dt I N E R C  
446 THATHAAYGARMGNTGY 462

QY 3 TCATTATGACGATGC 19

Ct I N E R C

RESULT 6 #type complete

ENTRY nuclear pore complex protein NSP16 - yeast (Saccharomyces cerevisiae)

ALTERNATE\_NAMES

nucleoporin NSP16; protein YN532.12c; protein YNR047c

ORGANISM

DATE

#formal\_name Saccharomyces cerevisiae  
07-May-1993 #sequence\_revision 07-May-1993 #text\_change  
06-Feb-1998

ACCESSIONS S28925; A44402; S52896; S28537

REFERENCE S28026

#authors Wimmer, C.; Doye, V.; Grandi, P.; Nehrbass, U.; Hurt, E.C.

#journal EMBO J. (1992) 11:5051-5061

#title A new subclass of nucleoporins that functionally interact

#accession S28925

#molecule\_type DNA

##residues 1-1113 #label WIM

##cross-references EMBL:X68108; NID:g4052; PID:g4053

REFERENCE A44402

#authors Wente, S.R.; Rout, M.P.; Blobel, G.

#journal J. Cell Biol. (1992) 119:705-723

#title A new family of yeast nuclear pore complex proteins.

#cross-references MUID:93054906

#accession A44402

#molecule\_type DNA

##residues 1-25; 'A', 27-535; 'G', 537-719; 'P', 721-1017; 'Y', 1019-1022,

'Y', 1024-1113 #label MEN

##cross-references EMBL:Z15036; NID:g4075; PID:g4076

REFERENCE S52885

#authors Ogden, C.; Bowman, S.

#submission submitted to the EMBL Data Library, February 1995

#accession S52896

#molecule\_type DNA

##residues 1-1113 #label ODE

##cross-references EMBL:Z48502; NID:g695715; PID:g695727; MIPS:YMR047C

GENETICS

#gene SGD:NDP116; NSP16

##cross-references SGD:S0004650; MIPS:YMR047C

KEYWORDS #map\_position 13R

SUMMARY #length 1113 #molecular-weight 116234 #checksum 1663

Query Match 75.8%; Score 72; DB 2; Length 1113;

Best Local Similarity 55.6%; Pred. No. 2.53e+01;

Matches 10; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Dt A S F I N E

Dt 2887 GCWMSNTTYTHAAYGAR 2904

Cp 19 GCATCGTTCATTATGAA 2

Ct A S F I N E

RESULT 7

ENTRY S73078

TITLE #type complete

ORGANISM hypothetical protein c0611 - Sulfolobus solfataricus

DATE #formal\_name Sulfolobus solfataricus

09-Oct-1997 #sequence\_revision 24-Oct-1997 #text\_change

17-Mar-1999

ACCESSIONS S73078

REFERENCE S73076

#authors Sengen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan,

C.C.Y.; Liu, Q.Y.; Penny, S.L.; Young, F.; Schenk, M.E.;

Gaasterland, T.; Doolittle, W.F.; Ragan, M.A.; Charlebois,

R.L.

Mol. Microbiol. (1996) 22:175-191

Organizational characteristics and information content of an

archaeal genome: 156 kb of sequence from Sulfolobus

solfataricus P2.

#cross-references MUID:97055432

#accession S73078

#molecule\_type DNA

##status nucleic acid sequence not shown; translation not shown

##residues 1-112 #label SEN

##cross-references EMBL:Y08256; NID:g1707679; PID:e284006; PID:g1707690

##experimental\_source strain P2

##note the nucleotide sequence was submitted to the EMBL Data

Library, September 1996

SUMMARY #length 112 #molecular-weight 12826 #checksum 5129

Query Match 74.7%; Score 71; DB 2; Length 112;

Best Local Similarity 63.2%; Pred. No. 3.64e+01;

Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Dt A T F I N E

Dt 133 GCNACNTTYTHAAYGARG 151

Cp 19 GCATCGTTCATTATGAG 1

Ct A S F I N E

RESULT 8

ENTRY A71560

TITLE #type complete

ORGANISM hypothetical protein CT079 - Chlamydia trachomatis (serotype

D, strain UW3/Cx)

DATE #formal\_name Chlamydia trachomatis

13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change

21-Nov-1998

ACCESSIONS A71560

REFERENCE A71570

#authors Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe,

R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov,

R.L.; Zhao, O.; Koonin, E.V.; Davis, R.W.

Science (1998) 282:754-759

Genome sequence of an obligate intracellular pathogen of

humans: Chlamydia trachomatis.

#cross-references MUID:99000809

#accession A71560

#molecule\_type DNA

##status preliminary

##residues 1-147 #label ARN

##cross-references GB:AE001282; GB:AE001273; NID:g3328466; PID:g3328474

##experimental\_source serotype D, strain UW-3/Cx

GENETICS

#gene CT079

SUMMARY #length 147 #molecular-weight 17134 #checksum 7530

Query Match 74.7%; Score 71; DB 2; Length 147;

Best Local Similarity 61.1%; Pred. No. 3.64e+01;

Matches 11; Conservative 5; Mismatches 2; Indels 0; Gaps 0;



Db F I K E R  
Dt 222 YTYATHAARGARMGNTG 239  
||:|:|:|:|:|:|:|:|:|  
QY 1 CTCATTAAATGACGATG 18  
Qt F I N E R

RESULT 9  
ENTRY H70212 #type complete  
TITLE hypothetical protein BBA48 - Lyme disease spirochete plasmid A/lps4

ORGANISM #formal\_name Borrelia burgdorferi #common\_name Lyme disease  
DATE 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 05-Jun-1998

ACCESSIONS H70212  
REFERENCE A70100

#authors Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Claydon, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Uitterback, T.; Watney, L.; McDonald, L.; Atiach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.  
#journal Nature (1997) 390:580-586  
#title Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.

#cross-references MIMD:98055943  
#accession H70212

#status preliminary; nucleic acid sequence not shown;  
translation not shown

#molecule\_type DNA  
#residues 1-162 #label KLE  
#cross-references GB:AF000790; NID:g2690224; PID:g2690279; TIGR:BBA48  
#experimental\_source strain B31

GENETICS  
#genome plasmid  
SUMMARY #length 162 #molecular-weight 19294 #checksum 2475

Query Match 74.7%; Score 71; DB 2; Length 162;  
Best Local Similarity 55.6%; Pred. No. 3.64e+01;  
Matches 10; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Db Y I N E R  
Dt 369 YTAATHAARGARMGNTG 386  
||:|:|:|:|:|:|:|:|:|  
QY 1 CTCATTAAATGACGATG 18  
Qt F I N E R

RESULT 10  
ENTRY S29910 #type complete  
TITLE hypothetical protein 4 - vaccinia virus  
ORGANISM #formal\_name vaccinia virus  
DATE 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Sep-1997

ACCESSIONS S29910  
REFERENCE S29907  
#authors Amegadzie, B.Y.  
#submission submitted to the EMBL Data Library, January 1991  
#accession S29910

#status preliminary  
#molecule\_type DNA

##residues 1-213 #label AME  
#cross-references EMBL:X57318; NID:g622239; PID:g622243  
SUMMARY #length 213 #molecular-weight 25037 #checksum 908

Query Match 74.7%; Score 71; DB 2; Length 213;  
Best Local Similarity 60.0%; Pred. No. 3.64e+01;  
Matches 9; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Db F I N E  
Dt 456 YTYATHAARGARMG 470  
||:|:|:|:|:|:|:|:|:|  
QY 1 CTCATTAAATGACG 15  
Qt F I N E

RESULT 11  
ENTRY I40172 #type complete  
TITLE orotidine-5'-phosphate decarboxylase (EC 4.1.1.23) - Bacillus caldolyticus

ORGANISM #formal\_name Bacillus caldolyticus  
DATE 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 05-Jun-1998

ACCESSIONS I40172; S34324  
REFERENCE I40166

#authors Ghim, S.Y.; Neuhaud, J.

#journal J. Bacteriol. (1994) 176:3698-3707  
#title The pyrimidine biosynthesis operon of the thermophile Bacillus caldolyticus includes genes for uracil phosphoribosyltransferase and uracil permease.

#cross-references MIMD:94265723  
#accession I40172  
#status preliminary; translated from GB/EMBL/DBJ  
#molecule\_type DNA  
#residues 1-244 #label RES

GENETICS  
#gene pyrF  
#start\_codon GTG

CLASSIFICATION  
#superfamily orotidine-5'-phosphate decarboxylase;  
#orotidine-5'-phosphate decarboxylase homology  
#carbon-carbon lyase  
KEYWORDS #length 244 #molecular-weight 26579 #checksum 7530

Query Match 74.7%; Score 71; DB 2; Length 244;  
Best Local Similarity 61.1%; Pred. No. 3.64e+01;  
Matches 11; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db F I K E R C  
Dt 499 YTYATHAARGARMGNTGY 516  
||:|:|:|:|:|:|:|:|:|  
QY 2 TTCATTAAATGACGATGC 19  
Qt F I N E R C

RESULT 12  
ENTRY C71703 #type complete  
TITLE hypothetical protein RP444 - Rickettsia prowazekii  
ORGANISM #formal\_name Rickettsia prowazekii  
DATE 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 21-Nov-1998

ACCESSIONS C71703  
REFERENCE A71630

#authors Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Slicheritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.; Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland, C.G.

#journal Nature (1998) 396:133-140  
#title The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

#accession C71703  
#status preliminary; nucleic acid sequence not shown;  
translation not shown

##residues 1-293 #label AND  
#cross-references GB:AJ235271; GB:AJ235269; NID:g3860788; PID:e1342745;  
SUMMARY #length 293 #molecular-weight 33724 #checksum 8486

Query Match 74.7%; Score 71; DB 2; Length 293;  
Best Local Similarity 60.0%; Pred. No. 3.64e+01;  
Matches 9; Conservative 6; Mismatches 0; Indels 0; Gaps 0;



```

FEATURE
26-64          #domain GAL4 zinc binuclear cluster homology #label
               GAL4\
452-468        #domain transmembrane #status predicted #label TM1\
485-501        #domain transmembrane #status predicted #label TM2\
520-536        #domain transmembrane #status predicted #label TM3
SUMMARY
#length 701 #molecular-weight 81273 #checksum 5832

Query Match          74.7%: Score 71; DB 2; Length 701;
Best Local Similarity 47.4%: Pred. No. 3,64e+01;
Matches 9; Conservative 7; Mismatches 3; Indels 0; Gaps 0

Dd      L I N E L C
Dt      930 YTTNATHAAGARNTNGY 948
Qy      1 CTTCAITNATGAACGATGC 19
Qt      F I N E R C

RESULT 15
ENTRY S46672 #type complete
TITLE hypothetical protein YHR182w - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein H9186.3
ORGANISM #formal name Saccharomyces cerevisiae
DATE 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 12-Dec-1997
ACCESSIONS S46672
REFERENCE S46671
#authors Macri, C.
#submission submitted to the EMBL Data Library, February 1994
#description The sequence of S. cerevisiae cosmid 9186.
#accession S46672
##molecule_type DNA
##residues 1-785 ##label MAC
##cross-references EMBL:U00028; NID:g458909; PID:g458911; MIPS:YHR182w
GENETICS
#map_position 8R
SUMMARY #length 785 #molecular-weight 90108 #checksum 2910

Query Match          74.7%: Score 71; DB 2; Length 785;
Best Local Similarity 60.0%: Pred. No. 3,64e+01;
Matches 9; Conservative 6; Mismatches 0; Indels 0; Gaps 0

Dd      F I N E
Dt      2082 YTTTATHAAGARNG 2096
Qy      1 CTTCAITNATGAACG 15
Qt      F I N E

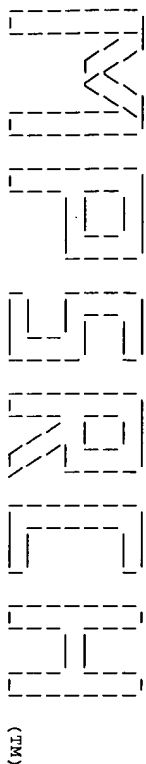
Search completed: Wed Nov 24 02:30:49 1999
Job time : 68 secs.

```



**THIS PAGE BLANK (USPTO)**





Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPerch\_ntp n.a. - n.a. Smith-Waterman search, using a protein database  
which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Nov 24 02:32:46 1999; MasPar time 4.78 Seconds  
Tabular output not generated. 674.763 Million cell updates/sec

Title: >US-09-103-287-5  
Description: (1-19) from US09103287.seq  
Perfect Score: 95  
N.A. Sequence: 1 CTCCTAATGACGATGC 19  
Comp: GAGCTAATGCTGCTACG

Scoring table: TABLE bkttranslated  
Gap 40

Mmatch STD : Dbase 0; Query 0

Searched: 77977 segs, 84804879 bases x 2

Post-Processing: Minimum Match 08  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 40.305; Variance 47.014; scale 0.857

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	76	80.0	592	1	Y036_HAELN HYPOTHEETICAL ABC TRANS	1.72e+00
2	72	75.8	437	1	MORC_STAUU UDP-N-ACETYLMURAMATE--	9.13e+00
3	72	75.8	1113	1	N116_YEAST NUCLEOPORIN NUP116/NSP	9.13e+00
4	71	74.7	213	1	AT12_VACCV PUTATIVE A-TYPE INCLUS	1.37e+01
5	71	74.7	244	1	DCOP_BACCL OROITINE 5'-PHOSPHATE	1.37e+01
6	71	74.7	425	1	PPA3_SCHPO PROBABLE DNA-DIRECTED	1.37e+01
7	71	74.7	485	1	DNA4_HUMAN FATTY ALDEHYDE DEHYDRO	1.37e+01
8	71	74.7	654	1	N05M_RHIST NADH-UBIQUINONE OXIDOR	1.37e+01
9	71	74.7	716	1	RRP2_IAT11 RNA-DIRECTED RNA POLYM	1.37e+01
10	71	74.7	785	1	YHYZ_YEAST HYPOTHEETICAL 90.1 KD P	1.37e+01
11	71	74.7	1030	1	SPAB_BACSU SUBTILIN BIOSYNTHESIS	1.37e+01
12	71	74.7	1096	1	RPOB_GUTH DNA-DIRECTED RNA POLYM	1.37e+01
13	71	74.7	148	1	NOLJ_RHIFR MODULATION PROTEIN NOL	2.03e+01
14	70	73.7	180	1	US10_VZVD GENE 64 OR 69 PROTEIN.	2.03e+01
15	70	73.7	308	1	YDDO_ECOLI HYPOTHEETICAL ABC TRANS	2.03e+01
16	70	73.7	493	1	HXX_PLAFA HEXOKINASE (BC 2.7.1.1	2.03e+01
17	70	73.7	549	1	YICE_ECOLI HYPOTHEETICAL 60.5 KD P	2.03e+01
18	70	73.7	899	1	PR06_YEAST PRE-MRNA SPLICING FACT	2.03e+01
19	70	73.7	899	1	PR06_YEAST PRE-MRNA SPLICING FACT	2.03e+01

20	70	73.7	1073	1	Y126_HUMAN HYPOTHEETICAL PROTEIN K	2.03e+01
21	70	73.7	3066	1	POLG_SERVG GENOME POLYPROTEIN [CO	2.03e+01
22	67	70.5	176	1	APT_BORBU ADENINE PHOSPHORIBOSYL	6.43e+01
23	67	70.5	278	1	HBP_GUTH POTENTIAL HEME-BINDING	6.43e+01
24	67	70.5	379	1	ACDS_CLOAB ACYL-COA DEHYDROGENASE	6.43e+01
25	67	70.5	405	1	CAR8_CANAL CANDIDAPRPSIN 8 PRECUR	6.43e+01
26	67	70.5	410	1	NEUS_CHICK NEUROSERPIN PRECURSOR	6.43e+01
27	67	70.5	413	1	INVB_ZYMO EXTRACELLULAR SUCRASE	6.43e+01
28	67	70.5	414	1	GAG2_DROME RETROVIRUS-RELATED GAG	6.43e+01
29	67	70.5	509	1	CPVL_BRARE CYTOCHROME P450 19A1 (	6.43e+01
30	67	70.5	518	1	CPVL_ORYLA CYTOCHROME P450 XIXA1	6.43e+01
31	67	70.5	986	1	EP1B_STAP EPIDERMIN BIOSYNTHESIS	6.43e+01
32	67	70.5	1065	1	RPOB_MARO DNA-DIRECTED RNA POLYM	6.43e+01
33	67	70.5	1103	1	RPOB_CYAPA DNA-DIRECTED RNA POLYM	6.43e+01
34	67	70.5	1302	1	RRPL_BTV10 RNA-DIRECTED RNA POLYM	6.43e+01
35	67	70.5	1379	1	RPOB_ODSI DNA-DIRECTED RNA POLYM	6.43e+01
36	67	70.5	1630	1	MSPI_PLAFK MEROZOITE SURFACE PROT	6.43e+01
37	67	70.5	1639	1	MSPI_PLAFW MEROZOITE SURFACE PROT	6.43e+01
38	67	70.5	2238	1	RRPL_BUNYW RNA POLYMERASE (EC 2.7	6.43e+01
39	66	69.5	118	1	YFEB_ECOLI HYPOTHEETICAL 13.6 KD P	9.32e+01
40	66	69.5	451	1	YJEB_YEAST HYPOTHEETICAL 52.1 KD P	9.32e+01
41	66	69.5	539	1	Y033_METUA HYPOTHEETICAL 82.0 KD P	9.32e+01
42	66	69.5	724	1	Y033_METUA HYPOTHEETICAL 82.0 KD P	9.32e+01
43	66	69.5	921	1	Y1B7_YEAST HYPOTHEETICAL 105.5 KD	9.32e+01
44	66	69.5	1384	1	Y19G_YEAST HYPOTHEETICAL 182.0 KD	9.32e+01
45	66	69.5	1967	1	YG50_YEAST PUTATIVE HELICASE YGR2	9.32e+01

# ALIGNMENTS

RESULT	1	STANDARD:	PRT:	592 AA.
ID	Y036_HAELN	057335; 005006;		
AC	057335; 005006;			
DT	01-NOV-1997 (REL. 35, CREATED)			
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	HYPOTHEETICAL ABC TRANSPORTER ATP-BINDING PROTEIN H10036.			
GN	H10036.			
OS	HAEMOPHILUS INFLUENZAE.			
OC	BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;			
OC	HAEMOPHILUS.			
RN	SEQUENCE FROM N.A.			
RP	STRAIN-RD / KW20;			
RC	MEDLINE: 95350630.			
RA	FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,			
RA	KERLAVAGE A.R., BILT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,			
RA	SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,			
RA	WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,			
RA	UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SADEK D.M., BRADON R.C.,			
RA	ELNE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEORGHEN N.S.M.,			
RA	GRENN C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,			
RA	VENTER J.C.;			
RT	Whole-genome random sequencing and assembly of Haemophilus			
RT	influenzae Rd.;			
RL	SCIENCE 269:496-512(1995).			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE			
CC	(POTENTIAL).			
CC	-1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY			
CC	(ABC TRANSPORTERS). STRONG, TO H. INFLUENZAE H1467 AND			
CC	M. TUBERCULOSIS MTCY1A11.24C.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
DR	EMBL, U32689; G1572962; -			
DR	TIGR, H10036; -			



DR PROSITE: P500211: ABC-TRANSPORTER; 1.

KM PFAM: PF00005: ABC-TRAN; 1.

KW INNER MEMBRANE.

FT TRANSMEM 12 32 POTENTIAL.

FT TRANSMEM 58 78 POTENTIAL.

FT TRANSMEM 102 122 POTENTIAL.

FT TRANSMEM 191 211 POTENTIAL.

FT TRANSMEM 214 234 POTENTIAL.

FT TRANSMEM 299 319 POTENTIAL.

FT NP BIND 424 431 ATP (POTENTIAL).

SEQUENCE 592 AA: 67903 MW: 78188BD3 CRC32;

Query Match Best Local Similarity 52.9%; Score 76; DB 1; Length 592; Pred. No. 1.72e+00; Mismatches 1; Indels 0; Gaps 0;

Matches 9; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db Dt 723 YTTTATHAAYGARGMNY 739

Qy 1 CTTCATTATGACGAT 17

Qt F I N E R

RESULT 2

ID MURC-STRAU STANDARD: PRT: 437 AA.

AC 031211;

DT 15-JUL-1998 (REL. 36, CREATED)

DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE UDP-N-ACETYLTRANSFERASE (EC 6.3.2.8) (UDP-N-ACETYLTRANSFERASE)

DE MURC.

OS STAPHYLOCOCCUS AUREUS.

OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;

OC STAPHYLOCOCCUS.

RN [1]

RP SEQUENCE FROM N.A.

RL LOWE A.M., DERESIEWICZ R.L.;

RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -1 FUNCTION: CELL WALL FORMATION (BY SIMILARITY).

CC -1 CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLTRANSFERASE + L-ALANINE -

CC -1 ADP + ORTHOPHOSPHATE + UDP-N-ACETYLTRANSFERASE + L-ALANINE.

CC -1 PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.

CC -1 SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).

CC -1 SIMILARITY: BELONGS TO THE MURDER FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

DR EMBL: AF034076; G2642659; -

KW PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;

KW ATP-BINDING.

FT NP BIND 108 114 ATP (POTENTIAL).

SEQUENCE 437 AA: 49176 MW: 6468582 CRC32;

Query Match Best Local Similarity 75.8%; Score 72; DB 1; Length 437; Pred. No. 9.13e+00; Mismatches 3; Indels 0; Gaps 0;

Matches 10; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db Dt 1177 GCNMSYNTATTHAAYGARG 1195

Cp 19 GCATCGTTCATTATGAG 1

Ct A S F I N E

RESULT 3

ID N116\_YEAST STANDARD: PRT: 1113 AA.

AC 02630;

DT 01-JUL-1993 (REL. 26, CREATED)

DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE NUCLEOPORIN NUP116/NSP116 (NUCLEAR PORE PROTEIN NUP116/NSP116).

GN NUP116 OR NSP116 OR YMR047C OR YMR532.12C.

OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).

OC SACCHAROMYCES CEREVISIAE; HEMIASCOCYCETES; SACCHAROMYCETALES;

OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYCETES.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 93054906.

RA WENTE S.R., ROUT M.P., BLOBEL G.;

RT "A new family of yeast nuclear pore complex proteins."

RT J. CELL BIOL. 119:705-723(1992).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE: 93099880.

RA WIMMER C., DOYE V., GRANDI P., NEHRBASS U., HURT E.C.;

RT "A new subclass of nucleoporins that functionally interact with

RT nuclear pore protein NSP1."

RT EMBO J. 11:5051-5061(1992).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-S288C / AB972;

RA ODELL C., BOWMAN S., RAJANDREAM M.A.;

RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -1 FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.

CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF

CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.

CC -1 SUBUNIT: INTERACTS WITH KAP95.

CC -1 SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.

CC -1 DOMAIN: CONTAINS G-L-F-G REPEATS.

CC -1 SIMILARITY: BELONGS TO THE GLEF FAMILY OF NUCLEOPORINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

DR EMBL: Z15036; G4076; -

DR EMBL: X68108; G4053; -

DR PIR: 248502; G695727; -

DR PIR: S28537; S28537.

DR PIR: S28925; S28925.

DR PIR: A4402; A4402.

DR SGD: L0001293; NUP116.

KW NUCLEAR PROTEIN; TRANSPORT; REPEAT.

FT DOMAIN 205 715

FT 37 X 6 AA APPROXIMATE REPEATS OF

FT G-L-F-G.

FT CONFLICT 26 26 G -> A (IN REF. 1).

FT CONFLICT 536 536 S -> G (IN REF. 1).

FT CONFLICT 720 720 S -> P (IN REF. 1).

FT CONFLICT 1018 1018 S -> Y (IN REF. 1).

FT CONFLICT 1023 1023 I -> Y (IN REF. 1).

SEQUENCE 1113 AA: 116234 MW: 28F496AF CRC32;

Query Match Best Local Similarity 75.8%; Score 72; DB 1; Length 1113; Pred. No. 9.13e+00; Mismatches 2; Indels 0; Gaps 0;

Matches 10; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Db Dt 2887 GCNMSYNTATTHAAYGARG 2904

Cp 19 GCATCGTTCATTATGAG 2

Ct A S F I N E

RESULT 4



ID AT12 VACC STANDARD; PRT; 213 AA.  
AC P24758;  
DT 01-MAR-1992 (REL. 21, CREATED)  
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)  
DE 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)  
DE PUTATIVE A-TYPE INCLUSION PROTEIN (PROTEIN A26).  
OS VACCINIA VIRUS (STRAIN WR).  
OC VIRUSES; DNA VIRUSES, NO RNA STAGE; POXYRIRDAE; CHORDOPOXYRIRMAE;  
ONTHOPOXYRIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 91310644.  
RA AMEGADZIE B.Y., AHN B.-Y., MOSS B.:  
RT Identification, sequence, and expression of the gene encoding a Mr  
35,000 subunit of the vaccinia virus DNA-dependent RNA polymerase.";  
RL J. BIOL. CHEM. 266:13712-13718(1991).  
CC -1- FUNCTION: MATURE VIRUSES ARE OCCCLUDED INTO THE ATI, AND IT HAS  
BEEN ASSUMED THAT SUCH BODIES PROTECT THE VIRUS DURING  
DISSEMINATION FROM ANIMAL TO ANIMAL.  
CC -1- A CHARACTERISTIC FEATURE OF ATI IS THE FORMATION OF LARGE MASSES  
WITH NO SURROUNDING MEMBRANES IN THE CYTOPLASM OF INFECTED CELLS.  
CC -1- SIMILARITY: TO COMPOX VIRUS A-TYPE INCLUSION PROTEIN C-TERMINUS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: M61187; G335786; -  
DR EMBL: X57318; G62243; -  
DR PIR: S29910; S29910.  
SQ SEQUENCE 213 AA; 25037 MW; EBDLF80B CRC32;  
-----  
Query Match 74.7%; Score 71; DB 1; Length 213;  
Best Local Similarity 60.0%; Pred. No. 1.37e+01;  
Matches 9; Conservative 6; Mismatches 0; Indels 0; Gaps 0;  
-----  
Db F I N E  
Dt 456 YTYATHAAGARGMG 470  
Qt 1 CTCATTATGACG 15  
F I N E  
-----  
RESULT 5  
ID DCOP BACCL STANDARD; PRT; 244 AA.  
AC P46355;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE OROITIDINE 5'-PHOSPHATE DECARBOXYLASE (EC 4.1.1.23) (OMP  
DECARBOXYLASE).  
GN PYR.  
OS BACILLUS CALDOLYTICUS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;  
OC BACILLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DSM 405;  
RA MEDLINE: 94282293.  
RX GHIM S.Y., NIELSEN P., NEUHARD J.:  
RT Molecular characterization of pyrimidine biosynthesis genes from the  
thermophilic Bacillus caldolyticus.";  
RL MICROBIOLOGY 140:479-491(1994).  
CC -1- CATALYTIC ACTIVITY: OROITIDINE-5'-PHOSPHATE -> UMP + CO(2).  
CC -1- PATWAY: SIXTH AND LAST STEP IN THE BIOSYNTHESIS OF PYRIMIDINES.  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE OMP DECARBOXYLASE FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: X73308; G580762; -  
DR PROSITE: PS00156; OMPDECASE; 1.  
DR PFAM: PF00215; OMPdecase; 1.  
KM PYRIMIDINE BIOSYNTHESIS; LYASE; DECARBOXYLASE.  
FT ACT\_SITE 61 61  
SQ SEQUENCE 244 AA; 26579 MW; 2188344B CRC32;  
-----  
Query Match 74.7%; Score 71; DB 1; Length 244;  
Best Local Similarity 61.1%; Pred. No. 1.37e+01;  
Matches 11; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
-----  
Db F I K E R C  
Dt 499 TTYATHAAGARGMGNTGX 516  
Qt 2 TTCATTATGACGATGC 19  
F I N E R C  
-----  
RESULT 6  
ID RPA3-SCHPO STANDARD; PRT; 425 AA.  
AC Q14086;  
DT 15-DEC-1998 (REL. 37, CREATED)  
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE PROBABLE DNA-DIRECTED RNA POLYMERASE I 49 KD POLYPEPTIDE (EC 2.7.7.6)  
DE (A49).  
GN SPAC2F3.03C.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;  
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;  
OC SCHIZOSACCHAROMYCES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972;  
RA OLIVER K., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WOOD V.:  
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
SUBSTRATES. RNA POLYMERASE A IS ESSENTIALLY USED TO TRANSCRIBE  
RIBOSOMAL DNA UNITS.  
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE +  
RNA(N).  
CC -1- SUBUNIT: RNA POLYMERASE I CONSISTS OF 14 DIFFERENT SUBUNITS  
(BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE FOUND IN  
EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA PRECURSOR,  
POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE III FOR 5S  
AND TRNA GENES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: Z99165; E1132708; -  
DR TRANSFERASE; DNA-DIRECTED RNA POLYMERASE; TRANSCRIPTION;  
KM NUCLEAR PROTEIN.  
SQ SEQUENCE 425 AA; 48066 MW; 0DB51994 CRC32;  
-----  
Query Match 74.7%; Score 71; DB 1; Length 425;  
Best Local Similarity 62.5%; Pred. No. 1.37e+01;  
Matches 10; Conservative 5; Mismatches 1; Indels 0; Gaps 0;



DB I N E R  
772 TTYATHAAGARGMT 787  
11:11:11:11:11:11  
QY 2 TTCATTATGACGAT 17  
OT F I N E R

RESULT 7  
ID DHA4\_HUMAN STANDARD; PRT; 485 AA.  
AC P51648; Q93011;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE FATTY ALDEHYDE DEHYDROGENASE (EC 1.2.1.3) (ALDEHYDE DEHYDROGENASE,  
DE MICROSOEAL) (CLASS 3).  
GN ALDH10 OR FALDH.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND VARIANTS SLS.  
RX MEDLINE: 96122039.  
RA DE LAURENZI V., ROGERS G.R., HAMROCK D.J., MAREKOV L.N.,  
STEINERT P.M., COMPTON J.G., MARKOVA N., RIZZO W.B.,  
RT "Sjogren-Larsson syndrome is caused by mutations in the fatty  
RT aldehyde dehydrogenase gene."  
RL NAT. GENET. 12:52-57(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA ROGERS G.R., MARKOVA N.G., DE LAURENZI V., RIZZO W.B., COMPTON J.G.,  
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE: 97224488.  
RA CHANG C., YOSHIDA A.;  
RT "Human fatty aldehyde dehydrogenase gene (ALDH10): organization and  
RT tissue-dependent expression."  
RL GENOMICS 40:80-85(1997).  
RN [4]  
RP VARIANT SLS SER-315.  
RX MEDLINE: 97398635.  
RA SILEN A., JAGELL S., WADELINUS C.;  
RT "A missense mutation in the FALDH gene identified in Sjogren-Larsson  
RT syndrome patients originating from the northern part of Sweden."  
RL HUM. GENET. 100:201-203(1997).  
RN [5]  
RP VARIANTS SLS R-106; W-226; S-315; L-365; R-412, & VARIANT N-245.  
RX MEDLINE: 99045377.  
RA SILEN A., ANTON-LAMPRECHT I., BRAUN-QUENTIN C., KRAUS C.S.,  
RA SAVLI B.S., AYUSO C., JAGELL S., KUESTER W., WADELINUS C.;  
RT "Spectrum of mutations and sequence variants in the FALDH gene in  
RT patients with Sjogren-Larsson syndrome."  
RL HUM. MUTAT. 12:377-384(1998).  
CC -1- FUNCTION: CATALYZES THE OXIDATION OF LONG-CHAIN ALIPHATIC  
CC ALDEHYDES TO FATTY ACIDS. ACTIVE ON A VARIETY OF SATURATED AND  
CC UNSATURATED ALIPHATIC ALDEHYDES BETWEEN 6 AND 24 CARBONS IN  
CC LENGTH.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE ENDOPLASMIC  
CC RETICULUM MEMBRANE.  
CC -1- DISEASE: DEFECTS IN FALDH ARE THE CAUSE OF SJOGREN-LARSSON  
CC SYNDROME (SLS). SLS IS AN AUTOSOMAL RECESSIVE NEUROCTANEOUS  
CC DISORDER CHARACTERIZED BY A COMBINATION OF SEVERE MENTAL  
CC RETARDATION, SPASTIC DI- OR TETRAPLEGIA AND CONGENITAL ICHTHYOSIS  
CC (INCREASED KERATINIZATION). ICHTHYOSIS IS USUALLY EVIDENT AT  
CC BIRTH. NEUROLOGIC SYMPTOMS APPEAR IN THE FIRST OR SECOND YEAR OF  
CC LIFE. MOST PATIENTS HAVE AN IQ LESS THAN 60. ADDITIONAL CLINICAL  
CC FEATURES INCLUDE GLISTENING WHITE SPOTS ON THE RETINA, SEIZURES,  
CC SHORT STATURE AND SPEECH DEFECTS.  
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: L47162; G1082036; -  
DR EMBL: U75296; G1666530; -  
DR EMBL: U75286; G1666530; JOINED.  
DR EMBL: U75287; G1666530; JOINED.  
DR EMBL: U75288; G1666530; JOINED.  
DR EMBL: U75289; G1666530; JOINED.  
DR EMBL: U75290; G1666530; JOINED.  
DR EMBL: U75291; G1666530; JOINED.  
DR EMBL: U75292; G1666530; JOINED.  
DR EMBL: U75293; G1666530; JOINED.  
DR EMBL: U75294; G1666530; JOINED.  
DR EMBL: U75295; G1666529; -  
DR EMBL: U75286; G1666529; JOINED.  
DR EMBL: U75287; G1666529; JOINED.  
DR EMBL: U75288; G1666529; JOINED.  
DR EMBL: U75289; G1666529; JOINED.  
DR EMBL: U75290; G1666529; JOINED.  
DR EMBL: U75291; G1666529; JOINED.  
DR EMBL: U75292; G1666529; JOINED.  
DR EMBL: U75293; G1666529; JOINED.  
DR EMBL: U75294; G1666529; JOINED.  
DR EMBL: U46689; G1870244; -  
DR MIM: 270200; -  
DR PROSITE: PS00070; ALDEHYDE\_DEHYDR\_CYS; 1.  
DR PROSITE: PS00687; ALDEHYDE\_DEHYDR\_GLU; 1.  
DR PFM: PF00171; aldehyd; 1.  
DR HSSP: P1883; IAD3.  
KW OXIDOREDUCTASE; NAD; ENDOPLASMIC RETICULUM; MICROsome; TRANSMEMBRANE;  
KW ALTERNATIVE SPLICING; DISEASE MUTATION; POLYMORPHISM.  
FT DOMAIN 1..463  
FT TRANSERM 464..480  
FT NP\_BIND 185..190  
FT ACT\_SITE 207..207  
FT ACT\_SITE 241..241  
FT ACT\_SITE 106..106  
FT VARIANT 214..214  
FT VARIANT 226..226  
FT VARIANT 245..245  
FT VARIANT 314..315  
FT VARIANT 315..315  
FT VARIANT 365..365  
FT VARIANT 412..412  
FT VARSPLIC 482..485  
SQ SEQUENCE 485 AA; 54847 MW; 87C524CF CRC32;  
Query Match 74.7%; Score 71; DB 1; Length 485;  
Best Local Similarity 60.0%; Pred. No. 1.37e+01;  
Matches 9; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

DB F I N E  
DE 1050 TTYATHAAGARGMT 1064  
QY 1 CTCATTATGACG 15  
OT F I N E  
RESULT 8  
ID NUSM\_RHIST STANDARD; PRT; 654 AA.  
AC P50367;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE NADH-BIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).  
GN NDS OR NADS.  
OS RHIZOPUS STOLONIFER (RHIZOPUS NIGRICANS).



OC MITOCHONDRION.  
OC EUKARYOTA; FUNGI; ZYGOMYCOTA; ZYGOMYCETES; MUCORALES; MUCORACEAE;  
OC RHIZOPUS.  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DAOM 148428:  
RA PAQUIN B., ROEMER I., WANG Z., LANG B.F.:  
RT "A robust fungal phylogeny using the mtchondrially encoded nad5  
protein sequence."  
RL CAN. J. BOT. 73:5180-5185(1995).  
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U17011; G604924; -  
CC DR PFAM: PF00361; oxidored\_g1; 1.  
CC DR PFAM: PF00662; oxidored\_g1.N; 1.  
CC KW OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION.  
SQ SEQUENCE 654 AA; 72818 MW; AD926718 CRC32;

Query Match 74.7%; Score 71; DB 1; Length 654;  
Best Local Similarity 63.2%; Pred. No. 1.37e+01;  
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db A P F I N E  
Dt 601 GCNCNTYATHAAYGARG 619  
Cp 19 GCATCGTCTATGATGAG 1  
Ct A S F I N E

RESULT 9  
ID RRP2\_IAT1 STANDARD; PRT; 716 AA.  
AC P13175;  
DT 01-JAN-1990 (REL. 13, CREATED)  
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)  
DE RNA-DIRECTED RNA POLYMERASE SUBUNIT P2 (EC 2.7.7.48) (P2 OR PA  
DE PROTEIN).  
OS INFLUENZA A VIRUS (STRAIN A/SWINE/TOM/15/30).  
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; ORTHOMYXOVIRIDAE;  
OC INFLUENZA VIRUS A AND B GROUP.  
OC [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 90021190.  
RA OKAZAKI K., KAWAKA Y., WEBSTER R.G.:  
RT "Evolutionary pathways of the PA genes of influenza A viruses."  
RL VIROLOGY 172:601-608(1989).  
CC -1- SUBUNIT: INFLUENZA RNA POLYMERASE IS COMPOSED OF THREE SUBUNITS:  
CC P1 (OR PB1), P2 (OR PA), AND P3 (OR PB2).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M26076; G325070; -  
CC DR PFAM: PF00603; Flu\_PA; 1.  
CC KW RNA-DIRECTED RNA POLYMERASE.  
SQ SEQUENCE 716 AA; 82566 MW; E1FDF31E CRC32;

Query Match 74.7%; Score 71; DB 1; Length 716;  
Best Local Similarity 60.0%; Pred. No. 1.37e+01;  
Matches 9; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Db F I N E  
Dt 156 YTYTATHAAYGARG 170  
Cp 1 CTTCAATTAATGAACG 15  
Ct F I N E

RESULT 10  
ID RRP2\_IATF STANDARD; PRT; 716 AA.  
AC P13177;  
DT 01-JAN-1990 (REL. 13, CREATED)  
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)  
DE RNA-DIRECTED RNA POLYMERASE SUBUNIT P2 (EC 2.7.7.48) (P2 OR PA  
DE PROTEIN).  
OS INFLUENZA A VIRUS (STRAIN A/SWINE/TENNESSEE/26/77).  
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; ORTHOMYXOVIRIDAE;  
OC INFLUENZA VIRUS A AND B GROUP.  
OC [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 90021190.  
RA OKAZAKI K., KAWAKA Y., WEBSTER R.G.:  
RT "Evolutionary pathways of the PA genes of influenza A viruses."  
RL VIROLOGY 172:601-608(1989).  
CC -1- SUBUNIT: INFLUENZA RNA POLYMERASE IS COMPOSED OF THREE SUBUNITS:  
CC P1 (OR PB1), P2 (OR PA), AND P3 (OR PB2).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M26077; G325080; -  
CC DR PFAM: PF00603; Flu\_PA; 1.  
CC KW RNA-DIRECTED RNA POLYMERASE.  
SQ SEQUENCE 716 AA; 82872 MW; EC309BD2 CRC32;

Query Match 74.7%; Score 71; DB 1; Length 716;  
Best Local Similarity 60.0%; Pred. No. 1.37e+01;  
Matches 9; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Db F I N E  
Dt 156 YTYTATHAAYGARG 170  
Cp 1 CTTCAATTAATGAACG 15  
Ct F I N E

RESULT 11  
ID YHY2\_YEAST STANDARD; PRT; 785 AA.  
AC P38870;  
DT 01-FEB-1995 (REL. 31, CREATED)  
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE HYPOTHEICAL 90.1 KD PROTEIN IN OYE2-GND1 INTERGENIC REGION.  
GN YHR182W.  
OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACHAROMYCETALES;  
OC SACHAROMYCETACEAE; SACHAROMYCES.  
OC [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-S288C / AB972;  
RX MEDLINE: 94376003.  
RA JOHNSON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,  
RA DU Z., FAVELLO A., FULTON L., GATUNG S., GEISEL C., KIRSTEN J.,  
RA KUCABA T., HILLIER L., JIER M., JOHNSON L., LANGSTON Y.,  
RA LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENZES S., MOUSER L.,  
RA NHAN M., RIEKIN L., RILES L., ST PETER H., TREVASKIS E., VAUGHAN K.,  
RA VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,



RA VAUDIN M.:  
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
 RT VII.";  
 RL SCIENCE 265:2077-2082(1994).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U00028; G458911; -  
 DR PIR: S46672; S46672.  
 KM HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 785 AA: 90108 MW: 03DCD892 CRC32;  
 DB 2082 YTTATTHAAYGARMG 2096  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE SUBUNITIN BIOSYNTHESIS PROTEIN SPAB.  
 GN SPAB OR SPAD OR SPAE.  
 OS BACILLUS SUBTILIS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;  
 OC BACILLUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 6633;  
 RX MEDLINE: 94161492.  
 RA GUTOWSKI-DCKEL Z., KLEIN C., SIEGERS K., BOHM K., HAMMELMANN M.,  
 RA ENTIAN K.D.;  
 RT "Growth phase-dependent regulation and membrane localization of Spab,  
 RT a protein involved in biosynthesis of the lantibiotic subtilin.";  
 RT APPL. ENVIRON. MICROBIOL. 60:1-11(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 6633;  
 RX MEDLINE: 94368094.  
 RA KLEIN C., ENTIAN K.D.;  
 RT "Genes involved in self-protection against the lantibiotic subtilin  
 RT produced by Bacillus subtilis ATCC 6633.";  
 RT APPL. ENVIRON. MICROBIOL. 60:2793-2801(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 6633;  
 RX MEDLINE: 93015727.  
 RA CHUNG Y.J., HANSEN J.N.;  
 RT "Determination of the sequence of spab and identification of a  
 RT promoter in the subtilin (spa) operon in Bacillus subtilis.";  
 RT J. BACTERIOL. 174:6699-6705(1992).  
 RN [4]  
 RP SEQUENCE OF 641-1030 FROM N.A.  
 RC STRAIN-ATCC 6633;  
 RX MEDLINE: 92171481.  
 RA KLEIN C., KALETTA C., SCHNEIL N., ENTIAN K.-D.;  
 RT "Analysis of genes involved in biosynthesis of the lantibiotic  
 RT subtilin.";  
 RT APPL. ENVIRON. MICROBIOL. 58:132-142(1992).

RN [5]  
 RP SEQUENCE OF 852-1030 FROM N.A.  
 RC STRAIN-ATCC 6633;  
 RX MEDLINE: 92138640.  
 RA CHUNG Y.J., GREEN M.T., HANSEN J.N.;  
 RT "The subtilin gene of Bacillus subtilis ATCC 6633 is encoded in an  
 RT operon that contains a homolog of the hemolysin B transport  
 RT protein.";  
 RL J. BACTERIOL. 174:1417-1422(1992).  
 CC -1- FUNCTION: INVOLVED IN THE POSTTRANSLATIONAL MODIFICATION OF THE  
 CC LANTIBIOTIC SUBTILIN.  
 CC -1- SUBCELLULAR LOCATION: POSSIBLY ASSOCIATED WITH, AND ANCHORED TO,  
 CC THE CYTOPLASMIC SIDE OF THE MEMBRANE.  
 CC -1- SIMILARITY: TO S.EPIDERMIDIS EPID AND L.LACTIS NISB.  
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
 CC FRAMESHIFT THAT PRODUCES TWO SEPARATE ORFS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L24075; G451879; -  
 DR EMBL: U09819; G595316; -  
 DR EMBL: M99263; G143564; ALT\_SEQ.  
 DR EMBL: M99263; G143565; ALT\_SEQ.  
 DR EMBL: M86869; G143714; -  
 DR EMBL: M83944; G143358; ALT\_INIT.  
 DR EMBL: U38418; G1109688; -  
 DR PIR: A45740; A45740.  
 DR PIR: A43935; A43935.  
 KM TRANSPORT.  
 FT CONFLICT 126 126 L -> F (IN REF. 3).  
 FT CONFLICT 261 261 D -> N (IN REF. 3).  
 FT CONFLICT 304 305 LA -> AG (IN REF. 3).  
 FT CONFLICT 344 344 H -> K (IN REF. 3).  
 FT CONFLICT 553 553 N -> I (IN REF. 3).  
 FT CONFLICT 584 584 F -> L (IN REF. 3).  
 FT CONFLICT 647 647 P -> A (IN REF. 3).  
 FT CONFLICT 730 730 E -> V (IN REF. 3).  
 SQ SEQUENCE 1030 AA: 120528 MW: 8310CC33 CRC32;  
 DB 543 YTTNATHAAYGARYNTGY 561  
 DT 15-DEC-1998 (REL. 37, CREATED)  
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6).  
 GN RPOB.  
 OS GUILLARDIA THETA (CRYPTOMONAS PH1).  
 OC CHLOROPHYTES.  
 OC EUKARYOTA; CRYPTOPHYTA; CRYPTOMONADACEAE; GUILLARDIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA DOUGLAS S.E., PENNY S.L.;  
 RT "The plastid genome from the cryptomonad alga, Guillardia theta:  
 RT complete sequence and conserved syteny groups confirm its common  
 RT ancestry with red algae.";



RL J. MOL. EVOL. 0-0-0(1998).  
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
CC SUBSTRATES.  
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +  
CC RNA(N).  
CC -1- SUBUNIT: IN CHLOROPLAST THE RNA POLYMERASE IS COMPOSED OF FOUR  
CC SUBUNITS: ALPHA, BETA, BETA', AND BETA".  
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
DR EMBL: AF041468: G3603015; -  
DR PROSITE: PS01166; RNA\_POL\_BETA: 1.  
DR TRANSCRIPTION: DNA-DIRECTED RNA POLYMERASE; CHLOROPLAST.  
SQ SEQUENCE 1096 AA: 123294 MW: 63f64cd7 CRC32;  
  
Query Match 74.7% Score 71; DB 1; Length 1096;  
Best Local Similarity 47.1%; Pred. No. 1.37e+01;  
Matches 8; Conservative 7; Mismatches 2; Indels 0; Gaps 0;  
  
Db L I N E R  
Dt 2001 YNTNTHAAGARGMNY 2017  
QY :|||:||||:|:  
Qt CTCATTATGACGAT 17  
F I N E R  
  
RESULT 14  
ID NOLJ RHIFR STANDARD; PRT: 148 AA.  
AC P12780:  
DT 01-OCT-1989 (REL. 12, CREATED)  
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE NODULATION PROTEIN NOLJ (HOST-INDUCIBLE PROTEIN B).  
GN NOLJ.  
OS RHIZOBIUM FREDDII.  
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;  
CC RHIZOBIACEAE; SINORHIZOBIUM.  
CC [1]  
CC SEQUENCE FROM N.A.  
CC RP MEDLINE: 88086864.  
CC RX SADOWSKY M.J., OLSON E.R., FOSTER V.E., KOSSIAK R.M., VERMA D.P.S.;  
CC "Two host-inducible genes of Rhizobium fredii and characterization of  
CC the inducing compound."  
CC RT J. BACTERIOL. 170:171-178(1988).  
CC [2]  
CC SEQUENCE FROM N.A.  
CC RP BOUNY-MILLS K.L., KOSSIAK R.M., TULLY R.E., PUEPKKE S.G.,  
CC LOHRKE S.M., SADOWSKY M.J.;  
CC "Induction of the Rhizobium fredii nod box-independent nodulation gene  
CC ngl requires a functional nodD gene."  
CC RT MOL. PLANT MICROBE INTERACT. 7:305-308(1994).  
CC -1- FUNCTION: INVOLVED IN EFFICIENCY OF SOYBEAN NODULATION AND IN  
CC NODULATION DELAY.  
CC -1- INDUCTION: BY PLANT 4',7-DIHYDROXY-ISOFLAVONE OR DERIVATIVES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
DR EMBL: M19019: G152249; -  
DR EMBL: L26967: G435107; -

DR PIR: B43663; B43663.  
KW NODULATION.  
SQ SEQUENCE 148 AA: 15791 MW: 0A6FAFEC CRC32;  
  
Query Match 73.7% Score 70; DB 1; Length 148;  
Best Local Similarity 66.7%; Pred. No. 2.03e+01;  
Matches 10; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
  
Db L M N D  
Dt 171 RYTNATGAGAYGC 185  
QY :|||:||||:|:  
Qt 5 ATTAATGACGATGC 19  
L M N D  
  
RESULT 15  
ID US10.VZVD STANDARD; PRT: 180 AA.  
AC P09311:  
DT 01-MAR-1989 (REL. 10, CREATED)  
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE GENE 64 OR 69 PROTEIN.  
GN 64 OR 69.  
OS VARICELLA-ZOSTER VIRUS (STRAIN DUMAS) (VZV).  
CC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;  
CC ALPHAHERPEVIRINAE; VARICELLOVIRUS.  
CC [1]  
CC SEQUENCE FROM N.A.  
CC RP MEDLINE: 86306557.  
CC RX DAVISON A.J., SCOTT J.E.;  
CC "The complete DNA sequence of varicella-zoster virus."  
CC RT J. GEN. VIROL. 67:1759-1816(1986).  
CC [2]  
CC SEQUENCE FROM N.A.  
CC RP MEDLINE: 85107111.  
CC RX DAVISON A.J., SCOTT J.E.;  
CC "DNA sequence of the major inverted repeat in the varicella-zoster  
CC virus genome."  
CC RT J. GEN. VIROL. 66:207-220(1985).  
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-1 US10,  
CC HSV-1 66, HSV-4 ORF3, AND VZV 64/69.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
DR EMBL: X04370: G60058; -  
DR EMBL: X04370: G60053; -  
DR EMBL: X02132: G59982; -  
DR PIR: C27345; WZB64.  
SQ SEQUENCE 180 AA: 19868 MW: B51DED2 CRC32;  
  
Query Match 73.7% Score 70; DB 1; Length 180;  
Best Local Similarity 68.8%; Pred. No. 2.03e+01;  
Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
  
Db L M N D  
Dt 89 CNYTNATGAGAYGC 104  
QY :|||:||||:|:  
Qt 4 CATTAATGACGATGC 19  
L M N D  
  
Search completed: Wed Nov 24 02:33:37 1999  
Job time : 51 secs.



**THIS PAGE BLANK (USPTO)**



\*\*\*\*\*  
Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
\*\*\*\*\*  
(TM)

Mserch\_nrp n.a. - n.a. Smith-Waterman search, using a protein database  
which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Nov 24 02:31:07 1999; Maspar time 9.78 Seconds  
636.263 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-103-287-5  
Description: (1-19) from US09103287.seq  
Perfect Score: 95  
N.A. Sequence: 1 CTTCAATTATGAACGATGC 19  
Comp: GAAGTAATTAATTGCTGAC

Scoring table: TABLE bkttranslate2  
Gap 40

Nmatch STD : Dbase 0; Query 0

Searched: 179066 seqs, 163739223 bases x 2

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database:

spremb19  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mnc 8:sp-oranelle  
9:sp-phase 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 38.972; Variance 46.490; scale 0.838

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	80	84.2	405	5	002302	T07D10.5 PROTEIN.	4.05e+01
2	76	80.0	181	5	046041	1-EVIDENCE-PREDICTED B	2.39e+00
3	76	80.0	265	11	P70408	CADHERIN 10 (T2-CADHER	2.39e+00
4	76	80.0	421	2	005296	HYPOTHEICAL 46.1 KD P	2.39e+00
5	76	80.0	517	2	050018	COSMID B1764.	2.39e+00
6	76	80.0	541	5	021484	MO3C11.4 PROTEIN.	2.39e+00
7	76	80.0	1211	3	000737	CHROMOSOME SEGREGATION	2.39e+00
8	72	75.8	398	1	058652	HYPOTHEICAL PROTEIN M	1.29e+01
9	72	75.8	636	10	P93604	RUST RESISTANCE KINASE	1.29e+01
10	72	75.8	2374	5	045317	F22G12.5 PROTEIN.	1.29e+01
11	72	75.8	2531	5	021980	SIMILAR TO TENASCIN.	1.29e+01
12	71	74.7	111	14	009666	POLYMERASE (FRAGMENT).	1.95e+01
13	71	74.7	112	1	P84988	ORF C06011.	1.95e+01
14	71	74.7	147	2	084082	HYPOTHEICAL 17.1 KD P	1.95e+01
15	71	74.7	153	5	017269	HYPOTHEICAL PROTEIN (	1.95e+01
16	71	74.7	162	2	050941	HYPOTHEICAL 19.3 KD P	1.95e+01

RESULT	ID	1	PRELIMINARY:	PRT:	405 AA.	ALIGNMENTS
AC	002302:	01-JUL-1997 (TREMBLERL. 04, CREATED)				INFLUENZA A VIRUS A/MT
DT	01-JUL-1997 (TREMBLERL. 04, LAST SEQUENCE UPDATE)					INFLUENZA A VIRUS STRA
DT	01-JAN-1999 (TREMBLERL. 09, LAST ANNOTATION UPDATE)					F15E11.2 PROTEIN.
DE	T07D10.5 PROTEIN.					ALCOHOL DEHYDROGENASE
GN	T07D10.5 OR H16D19.2					ALCOHOL DEHYDROGENASE
OS	CAENORHABDITIS ELEGANS.					COSMID C32E8.
OC	EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					POTATIVE 81.3 KD TRANS
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					POLYMERASE.
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					COMPLEMENT COMPONENT C
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					C06B8.7.
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					LICHENYGIN SYNTHETASE
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					F15E11.5 PROTEIN.
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					DIPETIDE TRANSPORT AT
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					GTP-BINDING PROTEIN HO
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					MAL3P1.3 PROTEIN.
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					UDP-GLUCOSE DEHYDROGEN
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					PR47E.
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					CYTOSOLIC P450 AROMAT
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					OVARIAN AROMATASE.
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					HYPOTHEICAL 90.1 KD P
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					GLUTAMATE RECEPTOR SUB
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					GLUTAMATE RECEPTOR DEL
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					GLUTAMATE RECEPTOR CHA
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					RNA-DIRECTED RNA POLYM
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					RNA-DIRECTED RNA POLYM
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					CORE PROTEIN.
OC	RHABDITIDA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					NEURONAL MYOSIN HEAVY

Query Match 84.2%; Score 80; DB 5; Length 405;



Best Local Similarity 52.6%; Pred. No. 4.05e-01;  
Matches 10; Conservative 7; Mismatches 2; Indels 0; Gaps 0;  
DB F N N E R C  
DC 1131 YTTAATGAAGTGTG 1149  
:||||:||||:||||:  
QY 1 CTTCAATATGAACGATGC 19  
QT F I N E R C

RESULT 2  
ID 046041 PRELIMINARY; PRT: 181 AA.  
AC 046041:  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DE 1-EVIDENCE-PREDICTED BY CONTENT.  
GN EG.10384.5  
OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYROIDIDEA;  
RN DROSOPHILIDAE; DROSOPHILA.  
[1]  
RP SEQUENCE FROM N.A.  
RA MURPHY L., HARRIS D., BARRELL B.;  
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA BENOS P.;  
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AL009193; F1202173;  
SQ SEQUENCE 181 AA; 20734 MW; CFB29EC7 CRC32;

Query Match 80.0%; Score 76; DB 5; Length 181;  
Best Local Similarity 52.6%; Pred. No. 2.39e+00;  
Matches 10; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

DB F I N E L C  
DC 321 YTTATGAAGTGTG 339  
:||||:||||:||||:  
QY 1 CTTCAATATGAACGATGC 19  
QT F I N E R C

RESULT 3  
ID P70408 PRELIMINARY; PRT: 265 AA.  
AC P70408:  
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE CADHERIN 10 (T2-CADHERIN) (FRAGMENT).  
GN CDH10.  
OS MUS MUSCULUS (MOUSE).  
OC EUMARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCUROGNATHI; MORIDAE; MORINAE; MUS.  
RN [1]  
RP SEQUENCE OF 117-203 FROM N.A.  
RC STRAIN-C57BL/6; TISSUE-TESTES;  
RX MEDLINE: 96211873.  
RA MUNRO S.B., DUCLOS A.J., JACKSON A.R., BAINES M.G., BLASCHUK O.W.;  
RT "Characterization of cadherins expressed by murine thymocytes";  
RL CELL. IMMUNOL. 169:309-312(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6; TISSUE-TESTES;  
RX MEDLINE: 97033837.  
RA MUNRO S.B., BLASCHUK O.W.;  
RT "A comprehensive survey of the cadherins expressed in the testes of fetal, immature, and adult mice utilizing the polymerase chain reaction";  
RL BIOL. REPROD. 55:822-827(1996).  
RN [3]  
RP SEQUENCE FROM N.A.

RC STRAIN-C57BL/6; TISSUE-TESTES;  
RA MUNRO S.B., BLASCHUK O.W.;  
RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U69137; G1546833;  
DR MGD: MGI:107436; CDH10.  
DR PRAM: PF01043; Cadherin\_C-term; 1.  
FT NON-TER 1;  
SQ SEQUENCE 265 AA; 29350 MW; 4845761C CRC32;

Query Match 80.0%; Score 76; DB 11; Length 265;  
Best Local Similarity 52.9%; Pred. No. 2.39e+00;  
Matches 9; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

DB F I N E R  
DC 567 YTTATGAAGTGTG 583  
:||||:||||:||||:  
QY 1 CTTCAATATGAACGAT 17  
QT F I N E R

RESULT 4  
ID 005296 PRELIMINARY; PRT: 421 AA.  
AC 005296:  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 46.1 KD PROTEIN.  
GN MTC1364.06C.  
OS MYCOBACTERIUM TUBERCULOSIS.  
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIAE;  
OC ACTINOMYCETALES; CORNBACTERIAE; MYCOBACTERIACEAE; MYCOBACTERIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA DEVLIN K., CHURCHER C.M.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA BARRELL B.G., RAJANDREAM M.A.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RX MEDLINE: 96181548.  
RA PHILIPP W.J., POULER S., EIGMEIER K., PASCOPELLA L.,  
RA BALASUBRAMANIAN V., HEYM B., BERCH S., BLOOM B.R., JACOBS W.R. JR.,  
RA COLE S.T.;  
RT "An integrated map of the genome of the tubercle bacillus,  
Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium  
leprae";  
RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).  
DR EMBL: Z93777; R311071;  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 421 AA; 46097 MW; 1DF89D9D CRC32;

Query Match 80.0%; Score 76; DB 2; Length 421;  
Best Local Similarity 61.1%; Pred. No. 2.39e+00;  
Matches 11; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

DB F I N E Q C  
DC 1045 YTTATGAAGTGTG 1062  
:||||:||||:||||:  
QY 2 TTTCAATATGAACGATGC 19  
QT F I N E R C

RESULT 5  
ID 050018 PRELIMINARY; PRT: 517 AA.  
AC 050018:  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)







Db 9 linvleqfhdnavvlfmagadlqklnayldklnknaf 46  
 OY 400 LINVLEQFDNAVVLFMGAGDIQKLNAYLDKLNKNAF 437

RESULT 6  
 ID. W24585 standard; Protein; 286 AA.  
 AC W24585;  
 DT 04-AUG-1997 (first entry)  
 DE H. pylori cytoplasmic protein, 11253.aa.  
 KW Transmembrane; cytoplasmic; cell envelope; flagella; transport;  
 KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;  
 KW activator; inhibitor; bacterial life cycle; vaccine; immune;  
 KW detection; antisense; inhibition.  
 OS Helicobacter pylori.  
 FH Key Location/Qualifiers  
 FT misc\_difference 2 /note="encoded by AAS"  
 FT misc\_difference 133 /note="encoded by TYA"  
 FT /note="encoded by TYA"  
 FT misc\_difference 166 /note="encoded by RCC"  
 FT W09719098-A1.  
 PD 29-MAY-1997.  
 PR 15-NOV-1996; U18542.  
 PR 17-NOV-1995; US-561469.  
 PA (ASTR ) ASTRA AB.  
 PI Smith DH;  
 DR WPI: 97-298052/27.  
 DR N-PSDB; T77043.  
 PT Helicobacter pylori nucleic acid sequences and related proteins -  
 PS used for diagnostics and therapeutics  
 PS Claim 18: Page 131, 235pp: English.  
 CC This sequence represents an H. pylori cytoplasmic protein involved in  
 CC outer membrane or cell wall biosynthesis. This sequence showed  
 CC homology to N-acetylmuramate-Alanine ligase.  
 CC Helicobacter pylori has been strongly linked to chronic gastritis and  
 CC duodenal ulcer disease. The nucleic acid sequences of the invention  
 CC are used to evaluate compounds, especially activators or inhibitors of  
 CC bacterial life cycle, for the ability to bind an H. pylori nucleic acid  
 CC sequence. The nucleic acid sequences, and corresponding proteins, are  
 CC also useful for generating vaccines for immunising subjects against H.  
 CC pylori or for use in detecting the presence of Helicobacter species in  
 CC a sample. Antisense nucleic acid sequences of these sequences are  
 CC used to inhibit expression of a gene from Helicobacter species. H.  
 CC pylori whole genomic DNA was isolated and nebulised to a median size of  
 CC 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique  
 CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are  
 CC complementary to the BstXI-cut PMPX vectors, while the overhang is not  
 CC self-complementary. Therefore the linkers will not concatemise nor  
 CC will the cut vector re-ligate itself easily. The linker-adaptor inserts  
 CC were ligated to each of the 20 PMPX vectors to construct a series of  
 CC shotgun subclone libraries. The purified DNA samples were then  
 CC sequenced.  
 CC Note: The ORF/protein reference number for this sequence was obtained  
 CC from the related specification, W09640893.  
 CC Sequence 286 AA:

Query Match 7.7%: Score 239; DB 22; Length 286;  
 Best Local Similarity 28.3%; Pred. NO. 6.71e-09;  
 Matches 63; Conservative 55; Mismatches 88; Indels 17; Gaps 13;

Db 10 flfnpfcavlpvtepehlydhlerffaykyflidhaqrvi-ykedpfiknyskd- 67  
 OY 158 FLFNPFCVAVLPVTEDPHDHYERF-DINDVFDVAFQEMAHVKNKGLIMGDDENHRLKTEADV 216  
 Db 68 aI--vI-ekkdly--niyylkdgpytsf-elknlgafllwgljehnatna-slaais 119  
 OY 217 PLYYYGFKSDSDIYAQNIQ-ITDKGTAFDVYVVGGEFYDHFSLSPQYGDHRTVLNALAVIAIS 275  
 Db 120 aldelnleelrnmxlfnkgikrtfdilqknnlilddyahhptelgxtlksariyanln 179  
 OY 276 YLEKLDVTNIKEALETFEGVKRRFNETTIANOVIVDDYAHHPREISATIDTARK--KYPH 333

Db 180 tqekliivgahkysrlmdnleefkcflehcdriilipyva 222  
 OY 334 K-E-VYAVFQPHTFSTRQAFINFEAES-LCKADRVFLCEIFGS 373

RESULT 7  
 ID W20102 standard; Protein; 286 AA.  
 AC W20102;  
 DT 29-JUL-1997 (first entry)  
 DE H. pylori cytoplasmic protein, 11253.aa.  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacterium; life cycle; activator;  
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
 KW diagnosis.  
 OS Helicobacter pylori.  
 FH Key Location/Qualifiers  
 FT misc\_difference 2 /label="Unknown"  
 FT /label="Unknown"  
 FT /note="encoded by AAS"  
 FT misc\_difference 133 /note="encoded by TYA"  
 FT /label="Unknown"  
 FT misc\_difference 166 /note="encoded by RCC"  
 FT W09640893-A1.  
 PD 19-DEC-1996.  
 PR 06-JUN-1996; 009122.  
 PR 07-JUN-1995; US-487032.  
 PR 01-APR-1996; US-630405.  
 PA (ASTR ) ASTRA AB.  
 PI Bergljndh OT, Smith D, Mellgaard BL;  
 DR WPI: 97-052306/05.  
 DR N-PSDB; T67723.  
 PT Helicobacter pylori nucleic acid sequences and related  
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
 PS infection, and to detect Helicobacter  
 PS Claim 61: Page 335; 1481pp: English.  
 CC This sequence represents a H. pylori cytoplasmic protein involved in  
 CC outer membrane or cell wall biosynthesis.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds.  
 CC useful as potential H. pylori life cycle activators or inhibitors.  
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the bacterial  
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
 CC and the predicted coding regions defined by computer evaluation. To  
 CC identify likely H. pylori antigens for vaccine development, the amino  
 CC acid sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having identified  
 CC and determined the sequences of interest, particular regions can be  
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
 CC production, e.g. in E. coli hosts.  
 CC Sequence 286 AA:

Query Match 7.7%: Score 239; DB 22; Length 286;  
 Best Local Similarity 28.3%; Pred. NO. 6.71e-09;  
 Matches 63; Conservative 55; Mismatches 88; Indels 17; Gaps 13;

Db 10 flfnpfcavlpvtepehlydhlerffaykyflidhaqrvi-ykedpfiknyskd- 67  
 OY 158 FLFNPFCVAVLPVTEDPHDHYERF-DINDVFDVAFQEMAHVKNKGLIMGDDENHRLKTEADV 216  
 Db 68 aI--vI-ekkdly--niyylkdgpytsf-elknlgafllwgljehnatna-slaais 119  
 OY 217 PLYYYGFKSDSDIYAQNIQ-ITDKGTAFDVYVVGGEFYDHFSLSPQYGDHRTVLNALAVIAIS 275  
 Db 120 aldelnleelrnmxlfnkgikrtfdilqknnlilddyahhptelgxtlksariyanln 179  
 OY 276 YLEKLDVTNIKEALETFEGVKRRFNETTIANOVIVDDYAHHPREISATIDTARK--KYPH 333  
 Db 180 tqekliivgahkysrlmdnleefkcflehcdriilipyva 222







[illegible]

DT 09-NOV-1998 (first entry)







**THIS PAGE BLANK (USPTO)**







CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
SQ SEQUENCE 446 AA; 47982 MW; 998066 CN;  
Query Match 4.7%; Score 147; DB 2; Length 446;  
Best Local Similarity 24.1%; Pred. No. 1.36e-02;  
Matches 74; Conservative 91; Mismatches 117; Indels 25; Gaps 19;  
Db 15 LGLAKSGESA-ARLDDKGAIVTVNDGKPFEDNPAAGSLLEEGKIVTGGHPLLEDEEF 73  
QY 7 VGIGSGMSSLAQIMHDLGHEVGQSDIENYV-FTEV-ALRNKGKILPFG-ANNI-KEDM 62  
Db 74 ALMKVNGIIPYNNPMIEKALAKRIPVLTVEE-LAYLISEAPIIIGITSGNGKTTTWMIGE 132  
QY 63 -VVIOGNAPFASHEIYRAHQKLDVSYNDPLGQIIDQYTSVAVTCAGHKSTTGLSH 121  
Db 133 VLTAAOGHGLISNGIGYPASQVAQIASDKDTLVNELSSFQLMGVQEFHPEIATVITNIMPT 192  
QY 122 VMNGDKTISFLIGD-G-TG-MGL--PESDYFAFEACEYR-RHFLSYKPDYAIMTNIDFD 174  
Db 193 HIDHGSFSEVAAKWNQKMTAADFLVNFNODLTSKTEATV-VPESTLEKVDGAYLE 251  
QY 175 HPDYFKDINDVDFAPQEAHNVKKG-ITAMGDDEHL-RKIEADVPITYYGFKSDSDIYQ 232  
Db 252 DGQLYFRG--EV-VMAA--NEIGVP--GSHVNEALATIAVAKLRGVNDQIKETLSAF 303  
QY 233 NIQTIDKGTADVYVDGFEYDHFSLSPQYGDHTVLANLAVIAISYLEKLDVTNKEALETF 292  
Db 304 GGVKRL 310  
QY 293 GGVKRRF 299  
RESULT 2  
ID US-08-665-435A-2 STANDARD: PRT: 450 AA.  
XX xxxxxx  
DE Sequence 2, Application US/08665435A  
XX  
CC Sequence 2, Application US/08665435A  
CC Patent No. 5681694  
CC GENERAL INFORMATION:  
CC APPLICANT: Skatrud, Paul  
CC APPLICANT: Peery, Robert  
CC APPLICANT: Hoskins, Joann  
CC APPLICANT: Wu, Chyun-Yeh Earnest  
CC TITLE OF INVENTION: Biosynthetic Gene Murd of Streptococcus  
CC NUMBER OF SEQUENCES: 3  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Eli Lilly and Company  
CC STREET: Lilly Corporate Center  
CC CITY: Indianapolis  
CC STATE: Indiana  
CC COUNTRY: US  
CC ZIP: 46285  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/665,435A  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Webster, Thomas D  
CC REGISTRATION NUMBER: 39,872  
CC REFERENCE/DOCKET NUMBER: X-9900

CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 317-276-3334  
CC TELEFAX: 317-276-3861  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 450 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 450 AA; 48579 MW; 1010380 CN;  
Query Match 4.6%; Score 143; DB 1; Length 450;  
Best Local Similarity 24.4%; Pred. No. 2.58e-02;  
Matches 76; Conservative 89; Mismatches 117; Indels 29; Gaps 23;  
Db 15 LGLAKSGESA-ARLDDKGAIVTVNDGKPFEDNPAAGSLLEEGKIVTGGHPLLEDEEF 73  
QY 7 VGIGSGMSSLAQIMHDLGHEVGQSDIENYV-FTEV-ALRNKGKILPFG-ANNI-KEDM 62  
Db 74 ALMKVNGIIPYNNPMIEKALAKRIPVLTVEE-LAYLISEAPIIIGITSGNGKTTTWMIGE 132  
QY 63 -VVIOGNAPFASHEIYRAHQKLDVSYNDPLGQIIDQYTSVAVTCAGHKSTTGLSH 121  
Db 133 VLTAAOGHGLISNGIGYPASQVAQIASDKDTLVNELSSFQLMGVQEFHPEIATVITNIMPT 192  
QY 122 VMNGDKTISFLIGD-G-TG-MGL--PESDYFAFEACEYR-RHFLSYKPDYAIMTNIDFD 174  
Db 193 HIDHGSFSEVAAKWNQKMTAADFLVNFNODLTSKTEATV-VPESTLEKVDG 251  
QY 175 HPDYFKDIND-VEP--AFQEAHNVKKGITAMGD-E-HL-RKIEADVPITYYGFKSDSD 228  
Db 252 AYLEDGQLYFRG--EV-VMAA--NEIGVP--GSHVNEALATIAVAKLRGVNDQIKET 303  
QY 229 IYAGNIQITDKGAFDYVDGFEYDHFSLSPQYGDHTVLANLAVIAISYLEKLDVTNKEA 288  
Db 304 LSAFGVYKRL 314  
QY 289 LETFGVYKRRF 299  
RESULT 3  
ID US-08-843-309-2 STANDARD: PRT: 450 AA.  
XX xxxxxx  
AC  
XX  
DE Sequence 2, Application US/08843309  
XX  
CC Sequence 2, Application US/08843309  
CC Patent No. 5834270  
CC GENERAL INFORMATION:  
CC APPLICANT: Skatrud, Paul  
CC APPLICANT: Peery, Robert  
CC APPLICANT: Hoskins, Joann  
CC APPLICANT: Wu, Chyun-Yeh Earnest  
CC TITLE OF INVENTION: Biosynthetic Gene Mur D of Streptococcus pneumoniae  
CC NUMBER OF SEQUENCES: 3  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Eli Lilly and Company  
CC STREET: Lilly Corporate Center  
CC CITY: Indianapolis  
CC STATE: Indiana  
CC COUNTRY: US  
CC ZIP: 46285  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/843,309  
CC FILING DATE:



```

CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Webster, Thomas D
CC REGISTRATION NUMBER: 39,872
CC REFERENCE/DOCKET NUMBER: X-9900
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 317-276-3334
CC TELEFAX: 317-276-3861
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 450 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 450 AA; 48579 MW; 1010380 CN;

Query Match 4.6%; Score 143; DB 2; Length 450;
Best Local Similarity 24.4%; Pred. No. 2,586-02;
Matches 76; Conservative 89; Mismatches 117; Indels 29; Gaps 23

Dd 15 LGLAKGESA-ARLDKGAIVYNDGKPFEDNPAQSLSEEGIKVITGGHPELDEEF 73
Qy 7 VGIKSGMSGLAQMHDHGHVEQSDIENYV-FTEV-ALRNKGKIKLPFG-ANNI-KEDM 62

Dd 74 ALMYNPGIPIYNNPIEALAKRIPIVLEVE-LAVLISEAPIIIGITSGNGKTTTWIGE 132
Qy 63 -VVIOGNFASHSHEIYVAHOLKLDVYSYNDFIQIIDQYTSVAVTGAHGKSTTGLSLH 121

Dd 133 VLTAAQGHLLSGNIGYPAQOVAQIASDKDPLVNELSSFOLMGVQEFHPETAIVTINIMPT 192
Qy 122 VMNGKRTKSFLLGD-G--TG-MG-LP-ESDFAFAPACRYR-RHLSYKPDVAIMTNIDFD 174

Dd 193 HIDYHGFSEYVAAKWNIQNKMTAADLVLFNODLAKDLTSKTEATV-VPESTLEKVDG 251
Qy 175 HPDYFKIND-VFD--AFQEMAHNVKGIIMAGDD-E-HL-RKTEADVPYYGFKXSDSD 228

Dd 252 AYLEGQLYFYG--EV-YMAA--NEIVP--GSHNVENALATIAVAKLRVYDQTKET 303
Qy 229 IYANQIQITDGTGFADVYVDEYDFHFLSPQYGDHTVLAALAVIAISYLEKLDVTNIRKEA 288

Dd 304 LSAFGVAKHRL 314
Qy 289 LETFGVAKRF 299

RESULT 4 STANDARD; PRT; 267 AA.
ID US-08-934-481-4
AC xxxxxx
DT
XX
XX
DE Sequence 4, Application US/08934481
XX
CC Sequence 4, Application US/08934481
CC Patent No. 5929045
CC GENERAL INFORMATION:
CC APPLICANT: Wallis, Nicholas G.
CC APPLICANT: Fuenyo, Joanna L.
CC APPLICANT: Lonetto, Michael A.
CC TITLE OF INVENTION: NOVEL MURD
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Dechert Price & Rhoads
CC STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CC City: Philadelphia
CC STATE: PA
CC COUNTRY: US
CC ZIP: 19103
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS

```

CC		SOFTWARE:	FASTSEQ FOR Windows Version 2.0
CC		CURRENT APPLICATION DATA:	
CC		APPLICATION NUMBER:	US/08/934,481
CC		FILING DATE:	
CC		CLASSIFICATION:	514
CC		PRIOR APPLICATION DATA:	
CC		APPLICATION NUMBER:	
CC		FILING DATE:	
CC		ATTORNEY/AGENT INFORMATION:	
CC		NAME:	Dickinson, Todd Q
CC		REGISTRATION NUMBER:	28, 354
CC		REFERENCE/DOCKET NUMBER:	GMI0070A
CC		TELECOMMUNICATION INFORMATION:	
CC		TELEPHONE:	215-994-2252
CC		TELEFAX:	215-994-2222
CC		TELEX:	
CC		INFORMATION FOR SEQ ID NO:	4:
CC		SEQUENCE CHARACTERISTICS:	
CC		LENGTH:	267 amino acids
CC		TYPE:	amino acid
CC		STRANDEDNESS:	single
CC		TOPOLOGY:	linear
CC		SEQUENCE	267 AA; 29079 MW; 348410 CN;
Db		Query Match	4.3%; Score 134; DB 2; Length 267;
Oy		Best Local Similarity	53.8%; Pred.No.1.08e-01;
		Matches 21; Conservative	7; Mismatches 11; Indels 0; Gaps 0;
93		GSNNVNALATIAVAKLRGVNDTKETLSAFGSKHRL	131
261		GDHTVLNLAIVAIISYLEKLDYNIKEALETFGGVKRRF	299
		::   :     ::          :	
RESULT	5		
ID	US-08-271-364A-7	STANDARD;	PRT: 1019 AA.
AC	xxxxxx		
DT			
DE			
Sequence 7,	Application US/08271364A		
Patent No.	5756334		
GENERAL INFORMATION:			
APPLICANT:	PERLER, FRANCINE B.		
APPLICANT:	SOUTHWORTH, MAURICE W.		
TITLE OF INVENTION:	RECOMBINANT THERMOSTABLE DNA POLYMERASE		
NUMBER OF SEQUENCES:	42		
CORRESPONDENCE ADDRESS:			
ADDRESSEE:	NEW ENGLAND BIOLABS, INC.		
STREET:	32 TOZER ROAD		
CITY:	BEVERLY		
STATE:	MASSACHUSETTS		
COUNTRY:	US		
ZIP:	01915		
COMPUTER READABLE FORM:			
MEDIUM TYPE:	Floppy disk		
COMPUTER:	IBM PC compatible		
OPERATING SYSTEM:	PC-DOS/MS-DOS		
SOFTWARE:	PatentIn Release #1.0, Version #1.25		
CURRENT APPLICATION DATA:			
APPLICATION NUMBER:	US/08/271,364A		
FILING DATE:	06-JUL-1994		
CLASSIFICATION:	435		
PRIOR APPLICATION DATA:			
APPLICATION NUMBER:	US 07/811,421		
FILING DATE:	18-DEC-1991		
PRIOR APPLICATION DATA:			
APPLICATION NUMBER:	US 07/666,340		
FILING DATE:	17-APR-1991		
PRIOR APPLICATION DATA:			



CC APPLICATION NUMBER: US 07/626,057  
CC FILING DATE: 11-DEC-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/513,994  
CC FILING DATE: 26-APR-1990  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: WILLIAMS, GREGORY D.  
CC REGISTRATION NUMBER: 30901  
CC REFERENCE/DOCKET NUMBER: NEB-101  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (508) 927-5054  
CC TELEFAX: (508) 927-1705  
CC INFORMATION FOR SEQ ID NO: 7:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1019 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: peptide  
CC SEQUENCE 1019 AA; 118920 MW; 5466116 CN;  
SQ  
  
Query Match 3.1%; Score 97; DB 2; Length 1019;  
Best Local Similarity 21.3%; Pred. No. 2.98e+01;  
Matches 13; Conservative 22; Mismatches 23; Indels 3; Gaps 3;  
  
Db 726 REYLVE-FNAVADVLSLMEPEELKEMRIGTRNGFRMGTFVDIDEDFAKLGGYVSEGSAR 784  
QY 364 RYFLCEIFGSIRENSGALTIDLD-KIGG-ASFINEDLINVLEQFDNAVLFMGAGDIQ 421  
Db 785 K 785  
QY 422 K 422  
  
RESULT 6  
ID US-08-222-715B-26 STANDARD; PRT; 1019 AA.  
XX xxxxxx  
DT  
XX  
DE Sequence 26, Application US/08222715B  
CC Sequence 26, Application US/08222715B  
CC Patent No. 5834285  
CC GENERAL INFORMATION:  
CC APPLICANT: Comb, Donald G.  
CC APPLICANT: Perlier, Francine  
CC APPLICANT: Kucera, Rebecca  
CC APPLICANT: Jack, William E.  
CC TITLE OF INVENTION: RECOMBINANT THERMOSTABLE DNA  
CC TITLE OF INVENTION: POLYMERASE FROM ARCHAEABACTERIA  
CC NUMBER OF SEQUENCES: 27  
CC CORRESPONDENCE ADDRESSES:  
CC ADDRESSEE: GREGORY D. WILLIAMS, NEW ENGLAND BIOLABS,  
CC ADDRESS: INC.  
CC STREET: 32 TOZER ROAD  
CC CITY: BEVERLY  
CC STATE: MASSACHUSETTS  
CC COUNTRY: US  
CC ZIP: 01915  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/222,715B  
CC FILING DATE: 04-APR-1994  
CC CLASSIFICATION: 536  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/167,238  
CC FILING DATE: 15-DEC-1993

CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/686,340  
CC FILING DATE: 17-APR-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/626,057  
CC FILING DATE: 11-DEC-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/513,994  
CC FILING DATE: 26-APR-1990  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Williams, Gregory D.  
CC REGISTRATION NUMBER: 30901  
CC REFERENCE/DOCKET NUMBER: NEB-054C3FC2  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (508) 927-5054  
CC TELEFAX: (508) 927-1705  
CC TELEX:  
CC INFORMATION FOR SEQ ID NO: 26:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1019 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: protein  
CC SEQUENCE 1019 AA; 118920 MW; 5466116 CN;  
SQ  
  
Query Match 3.1%; Score 97; DB 2; Length 1019;  
Best Local Similarity 21.3%; Pred. No. 2.98e+01;  
Matches 13; Conservative 22; Mismatches 23; Indels 3; Gaps 3;  
  
Db 726 REYLVE-FNAVADVLSLMEPEELKEMRIGTRNGFRMGTFVDIDEDFAKLGGYVSEGSAR 784  
QY 364 RYFLCEIFGSIRENSGALTIDLD-KIGG-ASFINEDLINVLEQFDNAVLFMGAGDIQ 421  
Db 785 K 785  
QY 422 K 422  
  
RESULT 7  
ID US-08-680-726A-54 STANDARD; PRT; 345 AA.  
XX xxxxxx  
DT  
XX  
DE Sequence 54, Application US/08680726A  
CC Sequence 54, Application US/08680726A  
CC Patent No. 5804197  
CC GENERAL INFORMATION:  
CC APPLICANT: Haanes, Elizabeth J.  
CC APPLICANT: Frank, Rexann S.  
CC TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES  
CC NUMBER OF SEQUENCES: 92  
CC CORRESPONDENCE ADDRESSES:  
CC ADDRESSEE: Sheridan Ross & McIntosh  
CC STREET: 1700 Lincoln Street, Suite 3500  
CC CITY: Denver  
CC STATE: Colorado  
CC COUNTRY: U.S.A.  
CC ZIP: 80203  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/680,726A  
CC FILING DATE: 12-JUL-1996  
CC CLASSIFICATION: 424  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Connell, Gary J.







```
CC INFORMATION FOR SEQ. ID NO:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 345 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC
```

CC	LENGTH:	345	amino acids
CC	TYPE:	amino acid	
CC	STRANDEDNESS:	single	
CC	TOPOLOGY:	linear	
CC	MOLECULE TYPE:	peptide	
CC	FRAGMENT TYPE:	internal	
SO	SEQUENCE	345: AA: 40609 MW: 660444	CN

Query Match 3.0%; Score 94; DB 1; Length 345;



	Query Match	3.0%;	Score 94;	DB 1;	Length 345;
	Best Local Similarity	21.8%;	Pred. No.	4.58+0;	
	Matches	36;	Conservative	45;	Mismatches 70;
					Indels 14;
					Gaps 13;
Db	135 TFLTDELGLVLAAPAPN-OG-QRRYIOENEFYDENVOLPRTNC-YFSK-EDKFE	190			
	:       : : :   ::::    : :	:	:   :	:	:    :

D <sub>b</sub>	27	VSSPTPKKEEFSIANENDI-PTANDSOEEDDGTFLQSLLHQFNASOVYTTTTIIIPM	85
	:	:  :   :	:
	:	:   :	:
Q <sub>y</sub>	212	I EADVP-IYYGFGNDSDDIYAQNQLIDKGAFVYVDGEFFHFLSPQGDHTVLN-AL	265
	:	:  :   :	:
D <sub>b</sub>	86	MVTITTYIKORD	97
		:  :   :	:
Q <sub>y</sub>	270	AVIAISTYEKLD	281



Mon Nov 29 10:16:54 1999

US-09-103-287-2.rai

Page 8

[illegible]

```

CC ADDRESS: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406-0939
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/94/9, 637
CC FILING DATE: 14-OCT-1997
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/028, 370
CC FILING DATE: 15-OCT-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Glimm, Edward R
CC REGISTRATION NUMBER: 38,891
CC REFERENCE/DOCKET NUMBER: P50566
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-4478
CC TELEFAX: 610-270-5090
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 496 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 496 AA; 56064 MW; 1299408 CN;
SQ
Query Match 3.0%; Score 94; DB 2; Length 496;
Best Local Similarity 20.3%; Pred. No. 4.56e+01;
Matches 24; Conservative 35; Mismatches 53; Indels 6; Gaps 6;
Db 312 TAAVDTMAVLTSQGVQFAANGSOVKFDGSLTI-YNDSDKMKLPPMVGDVYK-QV-N 368
: : : : : | : : : : : : : : : : : : : : : : : :
Oy 226 SDDIT-AQNITITDKGTAFDYVDGEFTDHLSPQYGDPHYLNALAVIAISYLEKLDVTN 284
: : : : : | : : : : : : : : : : : : : : : : : :
Db 369 SKPE-QHFTOPARYSEATLIK-TLENGVGARPSPTYAPTIETIQRYYVRLAKRFP 424
| : : : : | : : : : : : : : : : : : : : : : : :
Oy 285 IKALETGGVKKRRPNETTIANQVIYVDYAHHPREISATIDTARKKYPHKREVAVAFP 342

```











```

TITLE      The complete genome sequence of the gram-positive bacterium
JOURNAL    Nature 390 (6657), 249-256 (1997)
MEDLINE    98044033
REFERENCE  2 (bases 1 to 215640)
AUTHORS    Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
TITLE      Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
JOURNAL    Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
            adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
            68 89 48
FEATURES   location/Qualifiers
SOURCE     1. 215640
            /organism="Bacillus subtilis"
            /strain="168"
            /db_xref="taxon:1423"
            complement(8. .24)
            /gene="ytmK"
            complement(8. .1001)
            /gene="ytmM"
            complement(99. .1001)
            /gene="ytmK"
            /function="unknown"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB14906.1"
            /db_xref="PID:e1185801"
            /db_xref="PID:e1185801"
            /db_xref="GI:2635412"
            /db_xref="GI:2635412"
            /db_xref="SPRMBL:034430"
            /translation="MKRLVFAFTGLSGLQIDLSIGMAYGVTSLSLLAFGTPPAVAS
            ASVHLAEVYTAAGSVSHIKFQNVKQTVQDLYPGSIGAFAGAFSLQSGDVAKPY
            ISLFLLLGGLVLLRFLFOYKPALEKKAHPNKRKQSLPGYIAGFADTGGGNGPVY
            TPILSRKGLSPRKVYGVDTSEFAIYASATAGFLISGMEDVNMNLWVESLMAGIITA
            APIAAMLYOKFHPOLMGVLYGSEFIIVNARLLINEMIANTVHPLIYTAISALMSAV
            LFLVSKIGNRNVYTAADVHAKER"
            complement(1020. .2270)
            /gene="hipO"
            complement(1020. .2270)
            /gene="hipO"
            /gene="hipO"
            /EC_number="3.5.1.32"
            /codon_start=1
            /transl_table=11
            /product="hipurate hydrolase"
            /protein_id="CAB14907.1"
            /db_xref="PID:e1185802"
            /db_xref="PID:e1185802"
            /db_xref="GI:2635413"
            /db_xref="GI:2635413"
            /db_xref="SPRMBL:034980"
            /translation="MSLDYMRNIESYPYQTGNDILTLKEESNPVNLTLEKOLIGI
            RRLHOYPELSEKEETAFIKCKKEGQIIRPAIKTGYFADIAGESEGPATIALRA
            DIDLPIEEKTGLPYASKKRIHMACGHDFTALLGAFLKENQDSLKRIKILRO
            PAERAGATKTYIEDGOLDIDAVYGLNKKDIAVGTGLTKGPMNAVDKRFKEIES
            KGHAHALPHNGFDPLIGASQLVALQTTISRVNVPLOSAITLVKINGSTWNIPTD
            VVIGVYTFDSEVSNVQKQRFVAVTEQISAFSIKAVNKHSGPPPLCNDPEATGLV
            RDAAHKATLQVIDPASPAGEDFAVYLEHIGSPAFEGTGDHDMHHPAFTIDETAIL
            KASYFLVESAKRLDSNESKISD"
            complement(2288. .2980)
            /gene="tibr"
            complement(2288. .2980)
            /gene="tibr"
            /function="regulation of riboflavin biosynthesis genes"
            /note="alternate gene name: ytmK"
            /codon_start=1
            /transl_table=11
            /product="transcriptional regulator"
            /protein_id="CAB14908.1"
            /db_xref="PID:e1185803"

gene
CDS
            /db_xref="PID:g2635414"
            /db_xref="GI:2635414"
            /db_xref="SWISS-PROT:P94465"
            /translation="MTIAGVYVKGGLGRKIGSPPTANVAKIHGLRNGVYGLATVN
            HQHFLGYMNVGKPTVGSNLEKTEILETFPHFRITVGEKIECSLETFIRERDRDSLE
            SLIKIKDKDISCAKRFELIGMAPNKRESLSHQELNDPLCLCKNNDLYGNKRV
            YNVINDMNFPEFGITQVAVRYRITLISFLFKEDNPKVSKYIRGAGGLADKLRFTS
            SYVESEENILG"
            complement(3027. .4355)
            /gene="ytmJ"
            complement(3027. .4355)
            /gene="ytmJ"
            /function="unknown"
            /note="similar to nitrilotriacetate monoxygenase"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB14909.1"
            /db_xref="PID:e1185804"
            /db_xref="PID:e1185804"
            /db_xref="PID:g2635415"
            /db_xref="GI:2635415"
            /db_xref="SPRMBL:034974"
            /translation="MTRADPIQFGAMTHGVGCTDQWRHDPDPSASTNIEFYKKAQ
            TAEGLSPTFIADGLTISEKSLPHEPLNRPPTITLSALSVTKNIGLVGFTSFTIE
            PFTISRLMSLDHISGGAGNVLVTSQEGAAARHKSNDLPETHEREIQHQLDVVR
            GLMNSWEHDAFIHNKKTGFQFDQAKLRLNHNKGYFQVEGLNIGRSKQGPVYFOAG
            SSETGROFAAKNADAIPTHSNSLEETAKAFADYKSRADGRODSVRIPIGSIPIYA
            DTRERAKKRYREFALIPINAVYIARFPDDVDLSYPLDEPDDIGYKGNKFOST
            TDRIKREAKARNLTIREVAQDMARPRLLTGTPEPRAVSLETFTNNAADGFTIGSDI
            PGLIDAFVEKVIPILOERGIRQDYRGGLTRENGLDIPHOVSLSHSH"
            complement(4352. .4633)
            /gene="ytmI"
            complement(4352. .4633)
            /gene="ytmI"
            /function="unknown"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB14910.1"
            /db_xref="PID:e1185805"
            /db_xref="PID:e1185805"
            /db_xref="PID:g2635416"
            /db_xref="GI:2635416"
            /db_xref="SPRMBL:034639"
            /translation="MSDVNINIVMSKKGCSYCEEVKNYLNKGFPPONIDVSEKELR
            DILOVKYGVHVPVEIGEIRGNQYQITEIGIEIHLDLALANHAQIKREAKR"
            complement(4648. .5652)
            /gene="ytmO"
            complement(4648. .5652)
            /gene="ytmO"
            /function="unknown"
            /note="similar to hypothetical proteins"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB14911.1"
            /db_xref="PID:e1185806"
            /db_xref="PID:e1185806"
            /db_xref="PID:g2635417"
            /db_xref="GI:2635417"
            /db_xref="SPRMBL:034846"
            /translation="MTRLSTIDDSLIJEGETADTLOHTYKLAQMAEECGYHRPWAE
            HHNNDIETAGSAPEVLLGYLAATSKIRLASGGVNLQHSYKVAQDFHLLSALAPRI
            DLGKGAPGFOGLSTDALQAEYKRPVQPEKLEELHFPVDFPDDHRAALAPRQ
            VDRKPGIPLFGSTESAISAALKGISVFAFYIIGEEVEYKEARAFDAHLPGSEAE
            FHLAPVFAHTKEAEKHIVYSRESIKVVLKDGKRVGSGEQAQVLEAVTEPYDII
            VOKIGITAGTKEPAEELTRLSCYKINDYVITPPIKNAYEKOLSTOLDVAYIAKR"
            complement(5649. .6428)
            /gene="hisp"
            complement(5649. .6428)
            /gene="hisp"
            /function="histidine transport protein (ATP-binding
            protein)"
            /protein_id="CAB14912.1"
            /db_xref="PID:e1185807"

```



```

gene
CDS
        /db_xref="GI:2635418"
        /db_xref="SPTREMBL:O34900"
        /translation="MTETKIKHKOFGHNLKGLNLTVRKEVYVITIGPSGSKT
RCNLTLRPREGIISTIDKQINRPSPKSKPHMLRQZANVPOQYHLFPKHYIE
EGULTIARKMKQDAIYAVENELKVLQGLDNLNYPQLSGGQKQVIGTALALAIH
LTLDELPAALDPELGEVLEVMLEINVTGATIMLVTHMFEPARVSDOVVFMDEG
EQGTEPEVFRHRTKNDRTQFLRRVSPYLEPEPKHEPYI"
        complement(6425..7132)
        /gene="ytmM"
        complement(6425..7132)
        /gene="ytmM"
        /function="unknown"
        /note="similar to amino acid ABC transporter (pem28ase)"
        /codon_start=1
        /transl_table=1

```

	Query Match	53.7%	Score 2929	DB 17	Length 215640	
	Best Local Similarity	47.3%	Pred. No. 0.0b+00			
	Matches	606	Conservative	253	Mismatches 422	Indels 0
					Gaps	0
Db 49508	GTAGGCTCTCATATATTTTGGATATCTCTGCTCCATGGAAGATGAGAACGGCTTATC	49567				
Cp 1281	RTANCCRTTYGNARYYTTTGATATCTCNCNCNCOCATRAANARNACAGCHTTRC	1222				
Db 49568	ATGAGCTTTTAAACAGATGTCTCATCTTTCTTCAATCAGCCTTGCCATTAATGATTTTCC	49627				
Cp 1221	RAAYGYCARNMACRTDATTNARCTCYCTRTDATTAAANMNCNCNCDDATYTTTRC	1162				
Db 49628	CTGGAGTACACGAGTCTGACGCTTCCGGGATCTACAGGGGCTGAGCGAAATATCGGA	49687				
Cp 1161	DATNARRTCYTGDAATNGTANRAGCCNCMSWRITTYCNCKDATNSMNCBADAATTTCRA	1102				
Db 49688	TAAATACACAGTCTGGCGCCGCTCAGGCTTCTCTCAATTTGTCGAAGAACTGCTGGT	49747				
Cp 1101	NARARNANCKRKCRCNGCYTTRCAANRBSWYCNCCBPAVITRTTNARAANAGCTTGNT	1042				
Db 49748	CCGGTAAATGATGAGGCTGGGAATACCGGCAAAATTTCCGATCAGATCTCTGTCT	49807				
Cp 1041	NCKMSWRAANGTRFNGSGYTGRAANACNCNACYCTTTRTNGGTATVYTTNCK	982				
Db 49808	TGCCGCTCCGATTTGTCATTTTATTTCTGTGGAGATGGGATAGTCAATCAATGACAC	49867				
Cp 981	NGCNGTTRCDATFNGTNGNSMDATYTCCKKNGRTRGRTRNGRTARCTCAACATATAC	922				
Db 49868	TTGATCCCAAGCTCCTCTCATTAATCTGGGTTGAGCCCCCAAGAGATTTGAGAGC	49927				
Cp 921	YTGRTTNCDAATNGTNGTYCTRTTAANCKNKYTTNACNCCNCBRANGITYTCANRGC	862				
Db 49928	ATGCTTAATGATGCTGGAATCAATTTCTTCATTAATGAGTAACGAATGACCGCAATGA	49987				
Cp 861	YTCYTTDPATRTTNGTNACRTCANRATYTTTCANRRATNSWMDATNGCDATNACGNARNGC	802				
Db 49988	GTTTAATCATTTGTGTGTCGCCGTACGCGAGGATATAAAGCATATGAGACGATTTGCG	50047				
Cp 801	RTTANRANCGTRFGRTRCNCRTAATGNGSNWRARARTRCRTARAATTCGCCRTC	742				
Db 50048	GACAAGACATCAAAAGTGTGCCCTCGGTGTTTTCAGATGTTCTGCGGTGSAATTC	50107				
Cp 741	NACRTAANCRTRCRAANGNGTNCCTTTTCMGTDATVYGDATRTTYTNGCTAATATTC	682				
Db 50108	ATTTCCTCCCCCGGCATATATACACACGGGAGCTGCGAGATTTTNGCGAGATG	50167				
Cp 681	RTCSWRCYCTTBANCRTRATATATATATNGNACRCRCNGCYTDATYTTNCKNARRTG	622				
Db 50168	CTCATCGTCTCCGACGCAATATATGCTTTGTGTCCTGAAGACCATCTCTTGGAAAGC	50227				
Cp 621	YTCRTCRNCNCCGACGDAATATCCCTTYTATTCRTTRTNGCNCATYCTYTGRAANGC	562				
Db 50228	GTCAAATACATCGTCAATGCTTGAGAAATTAATCAGGGGATCAAAATCAATATTCGTAT	50287				

CP 561 RCTAANAACRCCRTDATTTCCTTTAAARTATCNGGRTGRTGRTAARTCATDATTGTGTCAT 502

Db 50288 AATGGCATAATCTGGCTGATAGCTTAAAGAATGGCGCGGATTTGCGACGCTTAAACAC 50347

CP 501 DATGCGCTAATCNGGTTTATTANSMNARRAARTCNCKNRATYTCGACGATCTCBAANGC 442

Db 50348 AAGATATCCGTGTTTTCATTTCCCGCGGCTGTTCGCTGCGGATTAAGAAGACGTGG 50407

CP 441 RAAARTATCNSWYTCNGGNARNOCCCATNCGTNGCTCNCDDATTAARAAASMGITTT 382

Db 50408 TTTTGGCGTTTGGATCAGTCAGACGACAGACCCGCTAGTCGAGCTTTTGGCTGTGCGCC 50467

CP 381 YTTTTCNCCRTTCAATNACRTGSMNARNANCCNGTNGTNSMGITTTTCGCTGACGNC 322

Db 50468 CGTAAACGGCAACTCGTGAATTTTTCATGTAGTCGCTTAAAACTATGATAACGAT 50527

CP 321 NGTNACNCCNACNSMNTPTATYATCTCDATDATYTCNCCNARRAARTCTRTTANSMNAC 262

Db 50528 CACGGGAATTCGCCACACATCCGCTTTTCTATTTAGATGCGGTGTGGGAATCGGT 50587

CP 261 NACRTCNARITTAATATYGTGTGNCNCCNACDATTYTCRTGSMNSMNGCTAANGC 202

Db 50588 TCCGCGCATGATCTGTCATGCCGGGTTTAAATTTTCCGCTAAGAAAGAAATCGAT 50647

CP 201 NCCYTGTATNACNCCATCTCTCTCTTTTATATRTTNGCNCRRANGAARATTTTAT 142

Db 50648 ATTTCTTTTTCACGCGCTCTTGGCGTAAATAATTTTTCGATATCCGATCTTGGAC 50707

CP 141 NCCYTCTTTNKNARNCCNACYTCCGTAANACRTATRTTCTDATTCTCNSMNCCTGNAC 82

Db 50708 AGTATATTCATTAATCAGATATTTTGGGCAACGAGCATACCGGTCCTTTATTC 50767

CP 81 YTCRGNCCNARRCTGCACTDATTYTGNGCNARNSMNSWNCATNCCSMNCCYTTDATNCC 22

Db 50768 AACAAATGATTAACAGCTCAT 50788

CP 21 NACRAARTGRTATNGTCAAT 1

RESULT 3

LOCUS AF008220 220060 bp DNA BCT 04-FEB-1998

DEFINITION Bacillus subtilis rmb-dnaB genomic region.

ACCESSION AF008220

KEYWORDS 92293135

VERSION AF008220.1 GI:2293135

KEYWORDS

SOURCE

ORGANISM

Bacillus subtilis.

Bacillus subtilis.

Eubacteria; Firmicutes; Low G+C gram-positive bacteria;

Bacillaceae; Bacillus.

1 (bases 1 to 7450)

Green,C.J., Stewart,G.C., Hollis,M.A., Vold,B.S. and Bott,R.F.

Nucleotide sequence of the Bacillus subtilis ribosomal RNA operon, rmb

Gene 37 (1-3), 261-266 (1985)

JOURNAL MEDLINE 86031361

REFERENCE 2 (bases 153210 to 153762)

AUTHORS Connors,M.J., Mason,J.M. and Setlow,P.

TITLE Cloning and nucleotide sequencing of genes for three small, acid-soluble proteins from Bacillus subtilis spores

JOURNAL MEDLINE 86195826

REFERENCE 3 (bases 213161 to 218473)

AUTHORS Ogasawara,N., Moriya,S., Mazza,P.G. and Yoshikawa,H.

TITLE Nucleotide sequence and organization of dnaB gene and neighbouring genes on the Bacillus subtilis chromosome

JOURNAL MEDLINE 87117549

REFERENCE 4 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 5 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 6 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 7 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 8 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 9 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 10 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 11 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 12 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 13 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 14 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 15 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 16 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 17 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 18 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 19 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 20 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 21 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 22 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 23 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 24 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 25 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 26 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 27 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 28 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 29 (bases 200404 to 201881)

AUTHORS Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.

TITLE Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis

JOURNAL MEDLINE 87117549

REFERENCE 30 (bases 200404 to 201881)

AUTHORS Seki,T., Yosh



MEDLINE REFERENCE AUTHORS TITLE JOURNAL MEDLINE REFERENCE AUTHORS TITLE	87250247 5 (bases 201234 to 203212) Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H. Nucleotide sequence of the Bacillus subtilis phoR gene J. Bacteriol. 170 (12), 5935-5938 (1988) 89053932 6 (bases 142232 to 144147) Grundy,F.J. and Henkin,T.M. Cloning and analysis of the Bacillus subtilis rpsD gene, encoding ribosomal protein S4 J. Bacteriol. 172 (11), 6372-6379 (1990) 91035248 7 (bases 133624 to 134990) Henkin,T.M., Grundy,F.J., Nicholson,W.L. and Chambliss,G.H. Catabolic repression of alpha-amylase gene expression in Bacillus subtilis involves a trans-acting gene product homologous to the Escherichia coli lacI and galP repressors Mol. Microbiol. 5 (3), 575-584 (1991) 91260441 8 (bases 140810 to 142610) Henkin,T.M., Glass,B.L. and Grundy,F.J. Analysis of the Bacillus subtilis tyrS gene: conservation of a regulatory sequence in multiple trnA synthetase genes J. Bacteriol. 174 (4), 1299-1306 (1992) 92138624 9 (bases 217570 to 220060) Putzer,H., Gendron,N. and Grunberg-Manago,M. Co-ordinate expression of the two threonyl-tRNA synthetase genes in Bacillus subtilis: control by transcriptional antitermination involving a conserved regulatory sequence EMBO J. 11 (8), 3117-3127 (1992) 92347349 10 (bases 134990 to 141290) Grundy,F.J., Waters,D.A., Takaya,T.Y. and Henkin,T.M. Identification of genes involved in utilization of acetate and acetoin in Bacillus subtilis Mol. Microbiol. 10 (2), 259-271 (1993) 95020526 11 (bases 162129 to 164080) Grundy,F.J., Waters,D.A., Allen,S.H. and Henkin,T.M. Regulation of the Bacillus subtilis acetate kinase gene by CopA J. Bacteriol. 175 (22), 7348-7355 (1993) 94042910 12 (bases 7009 to 15526) Kiel,J.A., Boels,J.M., Beidman,G. and Venema,G. Glycogen in Bacillus subtilis: molecular characterization of an operator encoding enzymes involved in glycogen biosynthesis and degradation Mol. Microbiol. 11 (1), 203-218 (1994) 94195107 13 (bases 196487 to 200620) Jin,S. and Sonenshein,A.L. Identification of two distinct Bacillus subtilis citrate synthase genes J. Bacteriol. 176 (15), 4669-4679 (1994) 94321340 14 (bases 16985 to 19588) Ade,A., Koide,H., Kohno,T. and Watabe,K. A Bacillus subtilis spore coat polypeptide gene, cots Microbiology 141 (Pt 6), 1433-1442 (1995) 95400496 15 (bases 131934 to 133970) Boloitin,A., Khazak,V., Stoyanova,N., Ratmanova,K., Yomantas,Y. and Kozlov,Y. Identical amino acid sequence of the ara(G) gene products of Bacillus subtilis 168 and B. subtilis Marburg strain Microbiology 141 (Pt 9), 2219-2222 (1995) 96118703 16 (bases 25258 to 31212) Rowland,B., Hill,K., Miller,P., Driscoll,J. and Taber,H. Structural organization of a Bacillus subtilis operon encoding menaquinone biosynthetic enzymes Gene 167 (1-2), 105-109 (1995)	MEDLINE REFERENCE AUTHORS TITLE JOURNAL MEDLINE REFERENCE AUTHORS TITLE	96144257 17 (bases 196487 to 200620) Jin,S., De Jesus-Berrios,M. and Sonenshein,A.L. A Bacillus subtilis maleate dehydrogenase gene J. Bacteriol. 178 (2), 560-563 (1996) 96134995 18 (bases 129888 to 132207) Varon,D., Brady,M.S. and Price,C.W. Bacillus subtilis operon under the dual control of the general stress transcription factor sigma B and the sporulation transcription factor sigma H Mol. Microbiol. 20 (2), 339-350 (1996) 96310371 19 (bases 81540 to 91690) Bower,S., Perkins,J.B., Vocum,R.R., Howitt,C.L., Rahaim,P. and Pero,J. Cloning, sequencing, and characterization of the Bacillus subtilis biotin biosynthetic operon J. Bacteriol. 178 (14), 4122-4130 (1996) 96312354 20 (bases 49093 to 51682) Vocum,R.R., Perkins,J.B., Howitt,C.L. and Pero,J. Cloning and characterization of the metE gene encoding S-adenosylmethionine synthetase from Bacillus subtilis J. Bacteriol. 178 (15), 4604-4610 (1996) 96345628 21 (bases 100760 to 102298) Kappes,R.M., Kempf,B. and Bremer,E. Three transport systems for the osmoprotectant glycine betaine operate in Bacillus subtilis: characterization of Opad J. Bacteriol. 178 (17), 5071-5079 (1996) 96359357 22 (bases 1 to 220060) Lapidus,A., Galleron,N., Sorokin,A. and Ehrlich,S.D. Sequencing and functional annotation of the Bacillus subtilis genes in the 200 kb rmb-dnaB region Microbiology 143 (Pt 11), 3431-3441 (1997) 98048467 23 (bases 1 to 220060) Lapidus,A., Galleron,N., Sorokin,A. and Ehrlich,D. Direct Submission Submitted (13-JUN-1997) Laboratoire de Genetique Microbienne, INRA, Domaine de Vilvert, Jouy-en-Josas cedex 78352, France Location/Qualifiers 1. 220060 /organism="Bacillus subtilis" /db_xref="taxon:1423" 485. 2034 /product="16S ribosomal RNA" 2203. 5129 /product="23S ribosomal RNA" 5185. 5300 /product="5S ribosomal RNA" 5322. 3597 /note="codon recognized: GTA" /product="trnA-Vai" 5430. 5504 /note="codon recognized: ACA" /product="trnA-Thr" 5543. 5618 /note="codon recognized: AAA" /product="trnA-Lys" 5629. 5715 /note="codon recognized: CTG" /product="trnA-Leu" 5721. 5795 /note="codon recognized: GGC" /product="trnA-Gly" 5810. 5895 /note="codon recognized: TTA" /product="trnA-Leu" 5905. 5981 /note="codon recognized: CGT"
---	--	---	--



trNA /product="tRNA-Arg"  
5997..6073  
/note="codon recognized: CCA"  
/product="tRNA-Phe"  
6079..6151  
/note="codon recognized: GCA"  
/product="tRNA-Ala"  
6172..6248  
/note="codon recognized: ATG"  
/product="tRNA-Met"  
6251..6327  
/note="codon recognized: ATG"  
/product="tRNA-Ile"  
6334..6425  
/note="codon recognized: TCA"  
/product="tRNA-Ser"  
6443..6519  
/note="codon recognized: ATG"  
/product="tRNA-Met"  
6531..6607  
/note="codon recognized: GAC"

Note: remainder of annotations omitted.

Query Match 53.7%; Score 2929; DB 18; Length 220060;  
Best Local Similarity 47.3%; Pred. No. 0.00e+00;  
Matches 606; Conservative 253; Mismatches 422; Indels 0; Gaps 0;

Db 129638 ATGACTGTTATCATCTTTGTTGGATTAAGAGACCGGATATAGTCCTGTCCTCCCAATA 129697  
1 ATGACNCAATATACATCTTTGTTGGATTAAGAGACCGGATATAGTCCTGTCCTCCCAATA 60  
Db 129698 CTTCATGATTAATGATATCTCTCCAGATGATGATGATGATGATGATGATGATGATGAT 129757  
61 ATGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120  
Db 129758 ACAGCGCTGAAAAAGAAATTAAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 129817  
121 GTGCGTGTGAAAAAGAAATTAAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180  
Db 129818 GGCATGACAGTATGCGCGGAGACGATCCAGACAGCAGATCCGATGATGATGATGATGATGAT 129877  
181 GAT 240  
Db 129878 ATGCTGAGGAGATTCGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 129937  
241 CAYCARTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
Db 129938 TTCAGAGTGTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 129997  
301 TAYACNMSGTGTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
Db 129998 CACGAT 130057  
361 CAYGAT 420  
Db 130058 AATGAAACAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 130117  
421 YTNCGNAGMSGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
Db 130118 TATCAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 130177  
481 TAYARCCGAT 540  
Db 130178 AGCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 130237  
541 GAT 600  
Db 130238 ATGCGCTGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 130297  
601 ATGCGTGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
Db 130298 TATGCGACGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 130357

661 TAYGNTTAAAGCAVMSGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
Db 130358 ACACTTTGAT 130417  
721 ACGNTTGGAT 780  
Db 130418 GCGCCACCAATGAT 130477  
781 GGNAGCAVACNGNTTAAAGCAVMSGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
Db 130478 GATTCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 130537  
841 GAT 900  
Db 130538 GAGAAGCAGCTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 130597  
901 GAT 960  
Db 130598 AAAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 130657  
961 WSNCGNACNATHGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
Db 130658 CAGCCTCATGAT 130717  
1021 CARCNCATGAT 1080  
Db 130718 GCGCCGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 130777  
1081 AARCGNAGMNGTNTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
Db 130778 CTGACGAT 130837  
1141 YTNACNATHGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
Db 130838 ACATCTGTTTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 130897  
1201 ATHAAGTNTNGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
Db 130898 CAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 130918  
1261 CARAARTNCARAAGCTAT 1281

RESULT 4  
LOCUS BACUNM. 2220 bp DNA BCT 26-MAR-1997  
DEFINITION Bacillus subtilis UDP-N-acetylmuramate-alanine ligase gene, partial cds, and 3 ORF's.  
ACCESSION L31845  
VERSION 9556013  
KEYWORDS L31845.1 GI:556013  
SOURCE UDP-N-acetylmuramate-L-alanine ligase.  
ORGANISM Bacillus subtilis  
Bacillus subtilis  
Bacteria; Firmicutes; Low G+C gram-positive bacteria;  
Bacillaceae; Bacillus.  
REFERENCE 1 (bases 1 to 2220)  
Varon, D., Brody, M.S. and Price, C.W.  
Bacillus subtilis operon under the dual control of the general stress transcription factor sigma B and the sporulation transcription factor sigma H  
Mol. Microbiol. 20 (2), 339-350 (1996)  
FEATURES  
source location/Qualifiers  
1..2320  
/organism="Bacillus subtilis"  
/strain="168"  
/sub\_species="Marburg"  
/db\_xref="taxon:1423"  
CDS  
1..1049  
/note="Identification based on sequence similarity to E. coli enzyme"  
/codon\_start=3



Query Match	Best Local Similarity	Matches	Score	DB	Length
45.1%	48.0%	491	2457	17	2320
Conservative	206	Mismatches	326	Indels	Gaps
9	GTGATACGTTATCATTAAGCTTTTTCAGCGACTACATGAAAAAATTCACGAGTGTGCCGTT	68			
259	GTGATACGTTATCATTAAGCTTTTTCAGCGACTACATGAAAAAATTCACGAGTGTGCCGTT	318			
69	ACGGCGCAGCAGCGCAAAACGTCGACTACGAGTGTGCCGTTTCAGCGACTACATGAAAAAATTCACGAGTGTGCCGTT	128			
319	ACGGCGCAGCAGCGCAAAACGTCGACTACGAGTGTGCCGTTTCAGCGACTACATGAAAAAATTCACGAGTGTGCCGTT	378			

Db	129	AAACCGAGCGCTTCTTAAATCGGAGCGGAAACAGCCGACGGGAATGAAACAGCGAATAC	188
Oy	379	AAARARCCNNSTNTTYTNATHGGNGVGGNACNGSNATGGGTYTNCNGAAMSNGATAY	438
Db	189	TTTGTTGTTAAACCGCGCAATACCCGCCCATTTGCTTAAGCTATCACCAGATATGCG	248
Oy	439	TTTCGNTTGAACGNTNGYGARATAYMMNMNCATYTTTYTINMSWTATAAACCGATATAGCN	498
Db	249	ATTATGACGAATATGTGATTTGATCACCTGATTACTTACTCTCAACGATTGACGATGATTT	308
Oy	499	ATHATGACNAAATYTHGATYTTTGAYCAYCCNGATYTTATYTAAGATATHAAGATGATNTTY	558
Db	309	GAGCCTTCCAAAGATAGGCTCTCAAGTCAACCAAGGATATATTTGCTGGGSGAGCAT	368
Oy	559	GAYCANTTTCARARATGGCCNCATYAAATYGNAPARAGNATHHGCNTGGGAGATGATG	618
Db	369	GAGCATCTGCCGGAATAATCCATGCGAAGCGTCCCGGTTGTATATTATGCGACGGGAGAA	428
Oy	619	GARCAATYTNNGNAPARATHGARCGNGATYTNCCNATHATYTTATYAGWTYTAAGAWSN	678
Db	429	AATGATTTTCAACCGCAAAACATCGTAAAAACACGGAAGGACACACTTTTGATGCTTT	488
Oy	679	GAYGAYATHTYACGNCARAAATACACARAAHACNGAYTAARGAGACGCTTGTGAYGATNTAY	738
Db	489	GTCGCAATACGTTCTATGATAGCTTTATCTTATTCCTGGGTCGGGACACCAATATGATTA	548
Oy	739	GTNCAATYGGNCAATTTTATGATCACTTTTYTINMSNCCNCAATATGAGATCAATACGNTNTN	798
Db	549	AACATCATGGCGGTCATATGGCTATGCCATTTATGAAGAATGTATCCAGCATATTAA	608
Oy	799	AAACNTTNGCNGTNAHGCNATHMSNTAYTNGARAAATYTGATGATGACNAAATATHAR	858
Db	609	CATGCTCTCAATTCCTTTGGGGCGCTCAAAAGCAGATTCATGAGAACACCTTGGGAT	668
Oy	859	GARCAANTNARACNTTYTGNGNGNTNAPRMNMNTTYAAAGARACNACNATHGCNAA	918
Db	669	CAATGGTGATGATGATGATGATGCCCATTCATCCGACAGAAATPAAAAGTACATGAGGCG	728
Oy	919	CARNTNATHGTNATYGAATYATGNCACATYCCCMNGNARATHMSGCAACNATHGATACN	978
Db	729	GCAAGACAGAAGTATCCTGATCGGAAATTTGTCGGGATATTCACGCTCATACATTTACC	788
Oy	979	GCNNGNAPARAPATYCCNCAATARGARGTNGCNGNTTYTTCARCCNCAVACNTYWSN	1038
Db	789	CGAGCGAGACAGTCTCTTACCAATTTTGACAAAGCCGTGAGCGGGCCGATGTGTGAT	848
Oy	1039	MGNACNARACNTTYTNNAYGARTTYGCNGARWSNNTNTGAARGCAGATMGCTNTTY	1098
Db	849	TTATGCAATTTTTCGGCTGACCGCGTGTAGATGCGGAAAGCTACAGCTGGGTGATC	908
Oy	1099	YTNNGYARATHTTYYGNSNATHHMGNGARAAAYMSNGNGCYTATACNATACAGATYTN	1158
Db	909	CAGGAAAAATTCATATGCAACGATGATTTGAAGAAGATGACACATCTGTTTTAAAGCT	968
Oy	1159	ATHGAYAPARATHGNGNGCNGMSNTTYATHAAYARGAYTATHAAGTNTYGARCAR	1218
Db	969	CATGATATAACCTTCTCATCTTACATGAGGAGACGAGAGATATCCAAAAATATATGAGACC	1028
Oy	1219	TTTAYATAYACGNTNGTNTTNTTATATGNGGNGGNGATYATHCACAARYTNCARAAACGN	1278
Db	1029	TAC 1031	
Oy	1279	TAT 1281	
RESULT	5		
LOCUS	POYMURC	2454 bp	DNA
DEFINITION	Porphyryomonas gingivalis gene for MurC, FlsQ, partial and complete cds		
ACCESSION	D84504	D28916	
NTID	g2780741		
VERSION	D84504.1	GI:2780741	







Cp	153	NARDAATTTDANCCYTRTNCNKAANGCACYTCGTGTAANAACRTARTTYTCDATPTGC	94
Dc	1364	YYY	1423
Cp	93	NSMWCCTGYGNACYYCRGNCNNARRCTCGCATDTATYGNCANRSMWSCATCNCSM	34
Dc	1424	YYYYYYYYYYYYGGTAGCAAAAT	1443
Cp	33	NCCTTTTADTNCCNACRAAART	14
RESULT	7	AEO01213	14838 bp DNA BCT 16-JUL-1998
LOCUS		Treponema pallidum section 29 of 87 of the complete genome.	
DEFINITION		AEO01213	AEO00520
ACCESSION		g3322606	
MID		AEO01213.1	GI:3322606
VERSION			
KEYWORDS			
SOURCE		Treponema pallidum.	
ORGANISM		Eubacteriales; Spirochaetales; Treponema.	
REFERENCE		1 (bases 1 to 14838)	
AUTHORS		Fraser,C.M., Norris,S.J., Weinstein,G.M., White,O., Sutton,G.G., Dodson,R., Gwin,M., Hickey,E.R., Clayton,R., Ketchum,K.A., Sodergren,E., Hardham,J.M., McLeod,M.P., Salzberg,S., Peterson,J., Khalak,H., Richardson,D., Howell,J.K., Chudamaram,M., Uterback,T., McDonald,L., Artlach,P., Bowman,C., Cotton,M.D., Fujii,C., Garland,S., Hatch,B., Horst,K., Roberts,K., Matthey,L., Weidman,J., Smith,H.O. and Venter,J.C.	
TITLE		Complete genome sequence of Treponema pallidum, the syphilis spirochete	
JOURNAL		Science 281 (5375), 375-388 (1998)	
MEDLINE		98332770	
REFERENCE		2 (bases 1 to 14838)	
AUTHORS		Fraser,C.M., Norris,S.J., Weinstein,G.M., White,O., Sutton,G.G., Dodson,R., Gwin,M., Hickey,E.R., Clayton,R., Ketchum,K.A., Sodergren,E., Hardham,J.M., McLeod,M.P., Salzberg,S., Peterson,J., Khalak,H., Richardson,D., Howell,J.K., Chudamaram,M., Uterback,T., McDonald,L., Artlach,P., Bowman,C., Cotton,M.D., Fujii,C., Garland,S., Hatch,B., Horst,K., Roberts,K., Matthey,L., Weidman,J., Smith,H.O. and Venter,J.C.	
TITLE		Direct Submission	
JOURNAL		Submitted (06-MAR-1998) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA	
FEATURES		Location/Qualifiers	
source		1..14838	/organism="Treponema pallidum"
gene		90..1814	/db_xref="taxon:160"
CDS		90..1814	/gene="TP0329"
		/note="similar to SP:P39148 PID:55686 GB:A1009126 percent identity: 48.52; identified by sequence similarity; putative"	
		/codon_start=1	
		/transl_table=11	
		/product="serine hydroxymethyltransferase (glyA)"	
		/protein_id="AAC65317.1"	
		/db_xref="PID:g3322607"	
		/db_xref="GI:3322607"	
		/transl_table=11	
		/product="cell division protein, putative"	
		/protein_id="AAC65318.1"	
		/db_xref="PID:g3322608"	
		/db_xref="GI:3322608"	
		/translation="MSGGQLEKNSIIISGLPWAQELSKKSGKYVNLFEVAGNIR DPLPHRDIGSFSPRINDYISEVFNGRNDITVPDKSAGLTFCLQEMLSAYLERMAR QIPTEALDFLSRDPKAPALYERFIIMNMKNRMVLIDYSLEVSEDIANLSET DRFCVTLMRAMANDPEFTMEDISVMLENENTDINSRTASSTVKHILPNENETFR REFLCYLTKEEILVERGERLTERIKGLTLVDLVHOLAETOTIDANKPDNTSR RLSHTSATLOFLEKRKOIIEHEGSLFEDTSCDVAVGVYVAKRYLNVRARI KGRADVLPMTYLISGPITCKSPFMVASFEGICPMRLCONFOATOGCINOSLERTK LNIKLALTVYAVMBADAVFSRRNAPGESIRPOLGPNTOHRNITWILTSTR PDLPLIDLRGRGADEHALAFYPETTKTELFEHLKKLIKADVAHPVIRIRRC AVSGSDLEILVRAOQLTAMERNRTWTODITRTIEDFIPPAYHEIELQNVAVLCBC TCKEMLPARYOHLDSKLVAADIREFLKLIMGEK"	
		complement(3649..3720)	
		/gene="tRNA-Gly-1"	
		/product="Gly ttc"	
		complement(3649..3720)	
		/gene="tRNA-Gly-1"	
		complement(3814..5280)	
		/gene="TP0331"	
		complement(3814..5280)	
		/note="similar to GB:L42023 SP:P43774 PID:1003990 PID:122490 PID:1204804 percent identity: 55.51; identified by sequence similarity; putative"	
		/codon_start=1	
		/transl_table=11	
		/product="phosphoglucanate dehydrogenase (gnd)"	
		/protein_id="AAC65319.1"	
		/db_xref="PID:g3322609"	
		/db_xref="GI:3322609"	
		/translation="MGADIGFIGLVMEENLYNTERGSEVAFNRFTTYVDFFLAG RAHGKRITSAHSIAELVSLARPRIIMLVAGSAVDVIOILLPLEKGVLVDIGN SHYQDTRIEMHAILEAGIHFGTGVSGEEGALRPSPISGSAOPMLPIPCIALA AKADGSTPCDWVSGDAGHYKMIHNEIGDMOIIIEGWFKMALGNSYEHNHTT FTRMNTGRHSYLEITTAAILAHODTDGPRILEKIIDAAOGKGTGRMVCALEEESR LTIITSWBASLSOKORACKAHVRPSPKYSKAETLSAOOREELVSALEDLYXA KIYSYKGFELLSHRKRGWILDERSIALSIMRGCTIRSFELKSIAFAPOCHDENK LVAPFAEBELRACGKWRTYAESVRLALPVPAALSALAFDFTGAALPANLLQAO RDYFGAHTYERTDAPRGEFFHTMTGTGTIAGTYSI"	
		5296..5421	
		/gene="TP0332"	
		5296..5421	
		/gene="TP0332"	
		/note="hypothetical protein; identified by Glimmer; putative"	
		/codon_start=1	
		/transl_table=11	
		/product="T. pallidum predicted coding region TP0332"	
		/protein_id="AAC65327.1"	
		/db_xref="PID:g3322617"	
		/db_xref="GI:3322617"	
		/translation="WRACSAFRFYLKGDGDLVLSVYTRYLCVLFEPVORSGSL" 5459..6151	
		/gene="TP0333"	
		5459..6151	
		/gene="TP0333"	
		/note="similar to GB:AEO00783 percent identity: 31.40; complement(1614..1696)	



D	14339	TCATGACGATTATGCATCATCGACTGCAATTAATAAAGCGCTCCGGGTAAAAA	14398
Oy	926	TGTTGATGATTAATCNCATCAICATCCNMGNGARATITWSGNCAKNTHTGAACNCGMNA	985
D	14399	CGTTCCTTCGCCGAAGAAGATTGTGTCGATTTTATGTATCCATACATATTCGCCGACCG	14458
Oy	986	AARAATATCAACCAYAAARARGTNNGCNGTWTTCARCNCATYACNTTYSNMGNACN	1045
D	14459	CAGCCCTCCACCGCAATTGGTGATGCTCTTCACAGATGCGGATGATGATTATTCGANG	14518
Oy	1046	ARGCTTTTNNANGARITTYGCGNMRMSVTTWTGYAARCCNCAIYMNGINTTYYTWTGTG	1105
D	14519	AGATTACGCTCTGCTCGGGANA	14541
Oy	1106	ARATHTTGYGMNSMATHMNGAR	1128
RESULT	8		
LOCUS	AE001180	15079 bp	DNA BCT 15-DEC-1997
DEFINITION	Borrelia burgdorferi	(section 66 of 70) of the complete genome.	
ACCESSION	AE001180	AE000783	
NID	9268875		
VERSION	AE001180.1	GI:2688755	
KEYWORDS			
SOURCE	Lyme disease spirochete.		
ORGANISM	Borrelia burgdorferi		
REFERENCE	Eubacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group.		
AUTHORS	1 (bases 1 to 15079) Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A., Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K., Gwinn,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D., Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J., Salzberg,S., Hanson,M., van-Vugt,R., Palmer,N., Adams,M.D., Goearney,J.D., Weidman,J., Uterback,T., Matthey,L., McDonald,L., Attiach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K., Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C. Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi Nature 390 (6660), 580-586 (1997)		
JOURNAL	2 (bases 1 to 15079) 98065943		
MEDLINE			
REFERENCE	Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A., Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K., Gwinn,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D., Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J., Salzberg,S., Hanson,M., van-Vugt,R., Palmer,N., Adams,M.D., Goearney,J.D., Weidman,J., Uterback,T., Matthey,L., McDonald,L., Attiach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K., Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C. Direct Submission Submitted (12-DEC-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA		
TITLE	Location/Ounalifiers		
JOURNAL	1. 15079		
FEATURES	/organism="Borrelia burgdorferi"		
SOURCE	/db_xref="taxon:139"		
	68. 244		
gene	/gene="BB0813"		
CDS	68. 244		
	/note="hypothetical protein; identified by Glimmer; putative"		
	/codon_start=1		
	/transl_table=1		
	/product="B. burgdorferi predicted coding region BB0813"		
	/protein_id="AAC67176.1"		
	/db_xref="PID:g2688771"		



gene 322. .1656  
/gene="BB0814"  
/note="similar to PID:606198 GB:U00096 PID:1789656 percent identity: 33.64; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="panthothenate permease (panf)"  
/protein\_id="AAC67168.1"  
/db\_xref="PID:92688763"  
/db\_xref="GI:2688763"  
/translation="MLNKYFLANNINFTYVALFFSSSYIASFFISGPSAVYYGL  
SFLLATIQIPITLIVFIIIVGQLNRSEKRIAINIDIRHRYESDPLALMSGFLI  
FPMFLISQOLIGAKLIEFWGIDYVGLFFPAFLVYVFGGFKAVAYDLDIGF  
LMLVSVIISFLKMDLGGGINLFTATSLDKSLSPNADLKPOYIISFWILIGV  
ILGPOIINNFIAFDENAIKFSPLSTFISFLVLMILIGFPAILLEPDLSPMDK  
VIANVALKYNPSCFEMFEGLISAIMSYDSMLLITSLSISITFIYKEDLKEDKIG  
RIMASNIFFIITILIFSLFPNFFLFIIFRFGALDEVSFFIIVFGLINVSXTIA  
FASMFGLIFLYISLFFGLINWFHHPVPSFFVSIFTELVNFFCKNSKVC"  
1666. .2559  
/gene="BB0815"  
1666. .2559  
/note="similar to GB:U02214 GB:I43867 SP:P47451 PID:406497  
PID:1045895 percent identity: 31.20; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAC67167.1"  
/db\_xref="PID:92688762"  
/db\_xref="GI:2688762"  
/translation="MRDKYIFLEVLANDRGRDLSILKILNFSKASIIKIRKGD  
RLNGKSHSCVCKGDKTYLYKSLAQNLTLDKQFNSIDPOYIRKRIIYEDDL  
VLDKRGITVHGKNSLDLYVSYLSQRLSRSPSVHLDNRTSIIIFANIN  
TARKLSEARSGSIIIRKTFAILLGEKSVVVKMLFRKRLKTFVEDKFNVAIT  
KNPLISCKRATLVEIIVETFTQIRSCSPNNHPLINDKRYCNKFKKSDYFLAFL  
VKNGTFFKSCFKPSLDF"  
2578. .3528  
/gene="BB0816"  
2578. .3528  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0816"  
/protein\_id="AAC67175.1"  
/db\_xref="PID:92688770"  
/db\_xref="GI:2688770"  
/translation="MMSMNFONSSDFELKILSKYKDPAISIKHKFVVKYVALVS  
AGTGSFRSHIVADKYSTPLITDGVILNNMKIIGSSKAFEDNFKAVRRVFEEDA  
HCRMLEVLAKEEFNKILITGLSKMIDISKLPLNFKITITDVSROETIKAR  
ISRQGEHVPAAPAEITISIRPNLLNSIKVEFKGWFARKNYIRSVRPHFEKG  
VLSIKRAVRAQIIEHCSEYEDRNIVYDILKIKDGNMYLFKFLNIPLGNNLNTTEM  
LRTYIITNVLKTIINILSIDIVIHKEFYDKDYLBESYEG"  
3518. .4924  
/gene="BB0817"  
3518. .4924  
/note="similar to GB:I42023 SP:P45066 PID:1006469  
PID:1221261 PID:1205384 percent identity: 28.70;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="UDP-N-acetylmuramate--alanine ligase (murC)"  
/protein\_id="AAC67166.1"  
/db\_xref="PID:92688761"  
/db\_xref="GI:2688761"  
/translation="MKVDPDDLNNITFVGIKSGACSLACFLNSKGYCEGVDSVSK  
YVDELISNNKISLYDNIYFSLKQDRSPDLVYSAYNRKDLQVLEKKEINIPILS  
YEPALGELSRKYISIGASHGKTTTATLGLVFNKLNPVIVGSSYKDKDSASAI  
ACISNFIYETCEYKXKHLNFSPLNMLITNVYDIEHDFPKNEALEEARLYINNLIK

gene 4935. .5801  
/gene="BB0818"  
4935. .5801  
/note="similar to GB:I10328 SP:P23839 GB:X14235 PID:290494  
PID:42721 percent identity: 29.96; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAC67165.1"  
/db\_xref="PID:92688760"  
/db\_xref="GI:2688760"  
/translation="MKSMTGFEEYEKIIIGNMYFSVNLKSYNGKLEPKFRPLPEISGY  
DLIDIRNLISKYISRGVFLNNGYKELVPSVNTIPNPIEASIRLDSLARTNLINIK  
ELSDGFSLKALIDEDSROEYIGLFKGYEEALHYNNGRSPGENTKSDIYS  
TLVIERDKTIVKACSDINVKLTASIKENYSKIMDERDRIINAEAKMAIRLDINE  
EIMRDSHIEFFYKNLEFETGKALEFISQEMHREITTMKNAVDLDIKNLINLKNL  
LEKIKQIRNE"  
5798. .6340  
/gene="BB0819"  
5798. .6340  
/note="similar to GB:L77117 SP:Q58071 PID:1591369 percent  
identity: 38.76; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="cytidylate kinase (cmk-2)"  
/protein\_id="AAC67164.1"  
/db\_xref="PID:92688759"  
/db\_xref="GI:2688759"  
/translation="MKILSGKSGGNTVSGMIARHGYEFINTFHDIAREHNIF  
SEFYKEKITGRNDYWDKYLDRLSVLSRKNNVYASLAIWISADLKIYIAKME  
VRAERIMTREGMSIDVLSFTIRBENDKRIALYINIDIDYSEIDLYDVNTINP  
NEVEFLRDEIDKRLKNS"  
6376. .6576  
/gene="BB0820"  
6376. .6576  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0820"  
/protein\_id="AAC67174.1"  
/db\_xref="PID:92688769"  
/db\_xref="GI:2688769"  
/translation="MTKVPKRIODFDGNYELVATIRTEQIIDNISLAHAFDO  
KILGAFMDVLTGKFRYSIEGR"  
6563. .7483  
/gene="BB0821"  
6563. .7483  
/note="similar to GP:216567 percent identity: 27.75;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="2-methylthio-N6-isopentyladenosine tRNA  
modification enzyme (miaA)"  
/protein\_id="AAC67163.1"  
/db\_xref="PID:92688758"  
/db\_xref="GI:2688758"  
/translation="MKEDRVYFTEGPAVAGSNILFHPKKAKELIIVDSIQYKEEN  
IASSKPSNKLKHKIHHLVDFLDPEKDYTGIFYEQALKIYKEIRQKKRIPFVGGA  
FYFKHLKGPSFTVPSYKIRIYVNNLELKGSKYLKELKLVNDPIRPNMLNKNDIYR  
IKRSLEVYQIGIPISQOKROSSFKFNVIIGIKRSPEDKTRISININMLNSGLL



Note: remainder of annotations omitted.

Query Match 5.8%; Score 318; DB 18; Length 15079;

Best Local Similarity 37.7%; Pred. No. 1.88e-13;

Matches 77; Conservative 44; Mismatches 83; Indels 0; Gaps 0;

Db 4542 TGGACGATTATGCTATCATCTAGGAAATTAATACTCTTTGGTATTAAAT 4601  
 Oy 929 TNGAYGATAYCNCAYCCCMGNGARATHMSNCNACNATGAYACNGCMGNAARA 988  
 Db 4602 TTTATAAGATAAAGGATATTTTGGATTATGCTTATACCTTTACACAAAG 4661  
 Oy 989 AATAACCCNCAVAAARGATNGTNGTTCATCCACNCAVACNTTYSNMGNCAARG 1048  
 Db 4662 AATTTTGGCCGATTGTTGAAGTTCTTAAGTTCGCCGATATATTTATTTGACAATA 4721  
 Oy 1049 CTTTYYTNAAYGARTTTCGNGRMSNTTGTGAAGCNGAIVNGNTTYYTNTGTGARA 1108  
 Db 4722 TATATCTTCAATAGGAAAT 4745  
 Oy 1109 THTTYGMSNATNMGNGARAAW 1132

RESULT 9  
 LOCUS SP086147 2268 bp DNA BCT 16-FEB-1997  
 DEFINITION *Synechococcus* PCC7942 UDP-N-acetylmuramate-alanine ligase (murC)  
 gene, partial cds, UDP-N-acetylenolpyruvylglucosamine reductase  
 (murB), and mercuric resistance operon regulatory protein (merB)  
 genes: complete cds.

ACCESSION U86147  
 MID 91841787  
 VERSION 086147.1 GI:1841787

KEYWORDS  
 ORGANISM *Synechococcus* PCC7942.

REFERENCE 1 (bases 1 to 2268)  
 Eubacteria; Cyanobacteria; Chroococcales; *Synechococcus*.

AUTHORS Kitzner, S., and Kaplan, A.  
 TITLE Cyanobacterial mutants capable of growing in the presence of high  
 concentration of zinc

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2268)  
 Kitzner, S., and Kaplan, A.

AUTHORS  
 TITLE Direct Submission  
 JOURNAL Submitted (22-JAN-1997) Plant Sciences, Hebrew University,  
 Jerusalem 91904, Israel

FEATURES  
 SOURCE location/Qualifiers  
 1..2268

/organism="Synechococcus PCC7942"

/strain="PCC7942"

/db\_xref="taxon:1140"

1..843

/gene="murC"

<1..843

/note="murC"

/note="similar to slr1423 in the cyanobase (the entire  
 genome sequence of *Synechocystis* sp. PCC 6803); the  
 translation of murC probably starts upstream of this  
 sequence"

/codon\_start=1

/transl\_table=1

/product="UDP-N-acetylmuramate-alanine ligase"

/protein\_id="AAB47522.1"

/db\_xref="PID:g1841788"

/db\_xref="GI:1841788"

/translation="PDHCDLEAVISFTKFTSDNCOLLANWDCPTVDRPLPTISYS  
 FDPARGADYTDVQFRSGSQARIMWEGELGRIHPLMEAHNLNALAIAACRH  
 GMDPASIREGLAFEGARREFRGSAQGIQFVDVAHPSLAAATLAALQIDSC  
 SRLPEPRALVAFOPHRSRTQAFLEAFSGFADLVISIDYAGERNPGLSG  
 TLADIAIQYANVYAPDLAEVQRLLQPDALFLGAGNLNQTIVPRLDHYAD  
 RSAAT 854..1741

854..1741

gene

/gene="murB"

CDS

854..1741

/gene="murB"

/note="similar to murB in *Escherichia coli* and similar to  
 a hypothetical protein in *Synechocystis* sp. PCC 6803  
 (Cyanobase N. slr1424)"

/codon\_start=1

/transl\_table=1

/product="UDP-N-acetylenolpyruvylglucosamine reductase"

/protein\_id="AAB47523.1"

/db\_xref="PID:g1841789"

/db\_xref="GI:1841789"

/translation="MTGAVSLLETLQOAVELAGFTSERVGLAOFDEPASVEAATA  
 WQARLDFPVTLEAGSNLIDRGLPLGLVLRLOGATFPLATGCVAAAGEPTP  
 RLMAAAROGSMGEMAVGIPGLGAVVNAAGOCMDILOSVEIRDOLETPS  
 GSVFROVHRSAVLOTGACVSPOLQDPFEESOVYTTSTNFRORRPPHPLPC  
 GIVRDFEVMHFRSELAG"

complement(1778..2212)

/gene="merR"

/complement(1778..2212)

/gene="merR"

/codon\_start=1

/transl\_table=1

/product="mercuric resistance operon regulatory protein"

/protein\_id="AAB47524.1"

/db\_xref="PID:g1841790"

/db\_xref="GI:1841790"

/translation="MKRIGELAKOVAVNGTIVYETLQIPSTGENGYRRYKOT  
 IODQIRKQOTIGSLDETRQILTYABETPESLVQTLNKLNSLEKLDQIDTF  
 KADLESYRDWQDTPHNGNEVRNLVNCNHSPLSAIEREE"

BASE COUNT 454 a 582 c 670 g 562 t

ORIGIN

Query Match 5.2%; Score 282; DB 17; Length 2268;

Best Local Similarity 38.2%; Pred. No. 4.74e-10;

Matches 63; Conservative 37; Mismatches 65; Indels 0; Gaps 0;

Db 509 TGTGGCTATTTTCAACCCCATGCTATAGCCGACCCACCAAGCATTTTGGCGAGTTTG 568  
 Oy 1007 TNGTNGCNGTTCATCARGCNCAYACNTTYSNMGNCAARGCTTYYTNAAYGARTTGG 1066  
 Db 569 CCCAGAGTTTGGCCCGACCTGCTCTAATCAGTATATCTATGACGCGGTGAGC 628  
 Oy 1067 CNGARMSNYTGTGAAGCNGAVMNGNTTYYTNTGAGARTTGTGMSNATNMGNG 1126  
 Db 629 GCATTCGCGGAGCTGACGCGTCAACTCTCGCCGATGCATGC 673  
 Oy 1127 ARAATWSNGNGCNTNACNATHCARGAYYTNATGAYARATHG 1171

RESULT 10  
 LOCUS AE000736 15862 bp DNA BCT 03-SEP-1998

DEFINITION *Aquifex aeolicus* section 68 of 109 of the complete genome.

ACCESSION AE000736 AE000657

MID 92983763

VERSION AE000736.1 GI:2983763

KEYWORDS

SOURCE

ORGANISM *Aquifex aeolicus*.

REFERENCE 1 (bases 1 to 15862)

AUTHORS Decker, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L.,  
 Graham, D.E., Overbeek, R., Sneed, M.A., Keller, M., Aujay, M.,  
 Huber, R., Feldman, R.A., Short, J.M., Olson, G.J., and Swanson, R.V.

THE COMPLETE GENOME OF THE HYPERTHERMOPHILIC BACTERIUM *Aquifex*  
*aeolicus*

JOURNAL Nature 392 (6674), 353-358 (1998)

MEDLINE 98196666

REFERENCE 2 (bases 1 to 15862)

AUTHORS Decker, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L.,  
 Graham, D.E., Overbeek, R., Sneed, M.A., Keller, M., Aujay, M.,  
 Huber, R., Feldman, R.A., Short, J.M., Olson, G.J., and Swanson, R.V.

THE COMPLETE GENOME OF THE HYPERTHERMOPHILIC BACTERIUM *Aquifex*  
*aeolicus*

JOURNAL Nature 392 (6674), 353-358 (1998)

MEDLINE 98196666

REFERENCE 2 (bases 1 to 15862)

AUTHORS Decker, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L.,  
 Graham, D.E., Overbeek, R., Sneed, M.A., Keller, M., Aujay, M.,  
 Huber, R., Feldman, R.A., Short, J.M., Olson, G.J., and Swanson, R.V.



JOURNAL Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego,  
CA 92121  
COMMENT Putative indicates no similarity to known proteins  
Hypoetical indicates similarity to a protein of unknown function.  
FEATURES  
Source location/Qualifiers  
1.15862  
/organism="Aquilifex aeolicus"  
/strain="VF5"  
/db\_xref="taxon:63363"  
233..937  
/gene="aq\_1345"  
233..937  
/gene="aq\_1345"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07332.1"  
/db\_xref="GI:2983773"  
/db\_xref="GI:2983773"  
/translation="MMGLKRLRLRVRAAHVYERLEDPKITEEGELBELKOLKEA  
MEALAOVKASQIFEREARENERAKALYKASLLLOAOGKISPEEAELKALILIO  
KELHEKNARLEEEAKROEMAVKIOAKIDELKQIARKEALTKLAKLARNARVK  
KNKQIAKYDPSDTIMLEMKKEVEEETLQAYVELAKQVALDYSEAEKKEVEY  
ESKQLEELKRRGLS"  
939..1469  
/gene="aq\_1347"  
939..1469  
/gene="aq\_1347"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07333.1"  
/db\_xref="PID:q2983774"  
/db\_xref="GI:2983774"  
/translation="MGFLERIKLFSKEEKEKQESLISNLEPKKGDITIDEGTWE  
VIDVALYDGASKKEKWEIRASRQFLSEEEKIYFEEIDPEELPEPGEHFRYRG  
KPEEYVYKGRRLYKAGKAKYIKNLESYPTIWEFRSEGDIMLEIWEDEYIAY  
KGRELOEMEIESILPR"  
1384..1953  
/gene="aq\_1348"  
1384..1953  
/gene="aq\_1348"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07334.1"  
/db\_xref="PID:q2983775"  
/db\_xref="GI:2983775"  
/translation="MRFGMSTKLKPTKGENYRNGKLRVFFVLSLLFLLSCSPKREVK  
GKPEEFKLNKYEKGIYRAILENKPODTEYCKVHLDLIKEGEYIILSPRYKEREK  
LSTRKICELFSKGRLEAVAGEPRTPSALKVTVFVELKGLPLFLEPRAOAGGGGCG  
ITFLPLIPMGCGYNEGRIRPTGRYK"  
1950..2756  
/gene="aq\_1349"  
1950..2756  
/gene="aq\_1349"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07335.1"  
/db\_xref="PID:q2983776"  
/db\_xref="GI:2983776"  
/translation="MSGFIATLFWYIFSKYVDPVLPFRKAKIRREIFGNKNLALSISY  
AGYFLGARSFYVYFESLFRFVLLIFVIFLLGVIIFDLIFLRKIDKKEILNG  
NAGAGITOGIYFLSILISASFWRKESFLSVISLILSLGMVLFSTLLMSRL  
KLNFEERKENSASIVLGSITLGVSVLYGAIISGEFMGSLIFDLFSTVLYVVSQV  
LWIVFYVVEFLFRKYLISSEYENNLASLITAFITLAVMG"  
2753..4234  
/gene="aq\_1350"  
2753..4234  
/gene="aq\_1350"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC07337.1"  
/db\_xref="PID:q2983778"  
/db\_xref="GI:2983778"  
/translation="MSVSNAIKICIEFTGFAGVAYESLTIATLYLGNVQMSVVI  
STFLSMGISHASRIIPDDKTPFLAVLEHLSLPLVPSVTAIHFANPDLHOTVI  
YGLSPVIGSLIGLEIPLAVRINNMTEELKVNISVLEKDYLGSVAGLLVAYLFLPK  
GLPLTAIILAGEFNLISAPLVRVLPKPKFLKFLAIFTEPLATVAGHKRITLYEOK  
FYGEETIHEEPYOKIVLTPRGKXSLYLDGLOESTLDEKRYHEHVLVHPASELKR  
YKALILGGGDLALRELKYPGEHLVADLPKMIPESSKNLVMRKINENSPFTRL  
KVESDAPNFYVKTKEKIDPVYVDLIDPTBSSARYSLEFYMSLKNKKEGDTITTO  
AGDTFKREVPFCISILTKKAGFVAPLVVYIPIFGMGVILGSKPELNFELKEK  
TEFLNREALAFYFTLGKSLDCEPNVEVNTLKLKLVLIYYKKION"  
4260..5306  
/note="aq\_1351"  
/gene="phoH"  
4260..5306  
/gene="phoH"  
/codon\_start=1  
/transl\_table=11  
/product="phosphate starvation-inducible protein"  
/protein\_id="AAC07324.1"  
/db\_xref="PID:q2983765"  
/db\_xref="GI:2983765"  
/translation="MEKVLAEERIDLGARYDERFVAIVGRDENIKFFSKFDPVRIYPR  
GTELIKGEDEKREVENFLKAVLKRLEHERFLAEVEKAKOFLKSAYKYLEOPG  
REKLEEEVYIILTRKKAIVPTEOKKVVYAIKNDIVIGAGCKEYLAAMA  
LOHKRGKYNKIILTRPAVEAGEKIGFLPTAIEKYNHPIITLYDHALYDMYIDKYN  
MLERNITEIAPLAFMRGRTLNDAFIILDEAQNATQDMKMFTRIGFGSKAVITGGVT  
QIDLPKKKESGLIEAMEVLEKNIIEGIEFFKOTDVVRHPIVAKIINAYOYERAKEOS  
SKTEESKESEGD"  
5230..5682  
/gene="aq\_1354"  
5230..5682  
/gene="aq\_1354"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC07338.1"  
/db\_xref="PID:q2983779"  
/db\_xref="GI:2983779"  
/translation="MSTRKQKNRYLVKLKRRKVRKDKIEKWAELASLALGNVELS  
VITDDQELRELNTYRKDKPDTVLSFPMGEGFGYKILGDVVISODIAREQARELG  
HSLPEEVKRLIYHGIYILGVDHEKGEKKKFRLEENYVLSKLKAL"  
complement(5649..6479)  
/note="aq\_1355"  
/gene="ltpA"  
complement(5649..6479)  
/codon\_start=1  
/transl\_table=11  
/product="lipoteic acid synthetase"  
/protein\_id="AAC07325.1"  
/db\_xref="PID:q2983766"  
/db\_xref="GI:2983766"  
/translation="MMKPVLRHSKLYEYKILKRSRLYNVCEESRCPNISRCFGKTA  
TMIILNRCRTRCARTCNVKEGFPKGVDPDEPRLRLBAVTKLGTIVITSVTRDLPD  
GGASHPAKCIKRVLENIEDIKVEVLIIPDRNGKRALVELLKERPAVLNHNVEVPLRY  
PSVRIGANKYRSLNITLWMSKEIDKSVYTKSAILGFERKEKEVIVMEDLRVSDDFL  
VIGQYQPSLKHHPVVKYSEEEFEFEIEGEMGFVYVSKPNARSYKAPESLST  
"  
6505..6939  
/gene="aq\_1356"  
6505..6939  
/gene="aq\_1356"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07336.1"  
/db\_xref="PID:q2983777"  
/db\_xref="GI:2983777"  
/translation="MGKVIOFPFGEPEKKEKELTEPTEDEFNQVLEVFKKDKLT  
LYPSQVDILGIVIGEKYGDKEFVVGILMWEHYKXSHSLVSSFAVILKGDLEKISF



gene  
LEAKNNYDMKLMVNTKYIKDKFHTTELMGLPELKERKPEKEI"  
7056.8462  
/note="ag\_1357"  
/gene="cyda"  
7056.8462  
/gene="cyda"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome oxidase d subunit I"  
/protein\_id="AAC07328.1"  
/db\_xref="PDB:2983769"  
/db\_xref="GI:2983769"  
/translation="MDAVLLARIQPLAMTHLEFPVITTLGSLLLAIITWRKREED  
DERHILIRIGEFTTLLEAINPAGVNVGIVLEEFEGTMSSEYKTVGDIFGAPLAE  
ALFEKFLVSEGLVLLIFGRDRVSPKVFVSTWMAVIGNLISALMILVANSQDPKPKF  
YLEGAALVLASFEDAVNVVSTPYRNVHFLAGVTGALFVAGISAVYILKKKHIDPVRK  
ESFRILAVIGMIALSLQIIAGDLHGCVARFQPKLAMEDKMETESGAGCIEYVSGE  
IPKLSILAVIHMDPAKVGIGDLVOTFOKIRKGEVGEDELPELMDLPKQAGOSIYV  
ITEENIPTWIDPEMSLRCGMVFLGFEFAITISLGFTFYODIIEGRNMLIKFESIPDL  
PIVAILIIMLVAVNGRQPVVYIILKTKDASPLPAQGLISILLESIIYFVLTIL  
GLMVQYKNGRRPEVA"  
8464.9564  
/note="ag\_1358"  
/gene="cydb"  
8464.9564  
/gene="cydb"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome oxidase d subunit II"  
/protein\_id="AAC07329.1"  
/db\_xref="PDB:2983770"  
/db\_xref="GI:2983770"

Note: remainder of annotations omitted.

Query Match	5.1%	Score 279	DB 18	Length 15862
Best Local Similarity	44.9%	Pred. No. 9,00e-10		
Matches 61	Conservative	26	Mismatches 49	Indels 0
			Gaps	0
Dp 10299	CGAATCTCCATAAAGTAATAGTCTTGTGATCGTGTGGGTGGAACACTACAGCAA	10358		
Cp 1066	CRAATTCCTTARRAANCYGTGNGNCNCKNSPRAAGCTGTGNGYTGRAANACGCAACNA	1007		
Dp 10359	GATTTTGTCCGGGTACATATCCCGAGTGAAGTTATTATACGGCCCTTTATCTCCGTAGGCT	10418		
Cp 1006	CYTCTTTGTGNGGRRATYTTTNCANGNGRTGATNGTNGCNSWDATYTCNCKNGSRT	947		
Dp 10419	GGTGTCCGTAATCGTC	10434		
Cp 946	GRTGNGCRTARTCRFC	931		
RESULT	11			
LOCUS	ECU67892	1998 bp	DNA	BCT
DEFINITION	Escherichia coli UDP-MurNac:L-alanine ligase (murC3) gene, complete cds.			11-JUN-1997
ACCESSION	U67892			
NID	92177093			
VERSION	U67892.1	GI:2177093		
KEYWORDS				
SOURCE	Escherichia coli.			
ORGANISM	Escherichia coli.			
REFERENCE	Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.			
AUTHORS	1 (bases 1 to 1998)			
TITLE	Matsuura, H., Matsushashi, M., Oka, A. and Sugino, Y.			
JOURNAL	Genetic and biochemical studies on cell wall peptidoglycan synthesis in Escherichia coli K-12			
MEDLINE	Biochem. Biophys. Res. Commun. 36 (4), 682-689 (1969)			
REMARK	69286442			
REFERENCE	Origin of mutant strain			
AUTHORS	2 (bases 1 to 1998)			
	Eveland, S.S., Pompliano, D.L. and Anderson, M.S.			

[illegible]



```
REFERENCE 1 (bases 1 to 2793)
AUTHORS Ikeda,M.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1990) Ikeda M., Institute of Applied
Microbiology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku,
Tokyo, 113, Japan
2 (bases 1 to 2793)
AUTHORS Ikeda,M., Wachi,M., Jung,H.K., Ishino,F. and Matsubashi,M.
TITLE Nucleotide sequence involving murG and murC in the mra gene cluster
region of Escherichia coli
JOURNAL Nucleic Acids Res. 18 (13), 4014 (1990)
MEDLINE 90326550
FEATURES
source location/Qualifiers
1..2793
/organism="Escherichia coli"
/db_xref="taxon:562"
/map="2mins"
<1..85
/codon_start=2
/transl_table=11
/product="f1em protein (partial) (27 AA)"
/protein_id="CA36866.1"
/db_xref="pid:809884"
/db_xref="gi:809884"
/db_xref="swiss-prot:P16457"
/translation="STAIIMLRIDYETRLERKQAFVRSR"
70..75
/feature="put. ribosome binding site"
82..1149
/codon_start=1
/transl_table=11
/product="murG protein (AA 1-355)"
/protein_id="CA36867.1"
/db_xref="pid:842055"
/db_xref="gi:42055"
/db_xref="swiss-prot:P17443"
/translation="MSGQRLMVMAGGSGHVPGLAVAHHLMAQMGVRLTGADR
MEADIVPKHIGIDIFIRISGLKGIKALIAPIRFNMRQARIMKAYKEDVYGM
GGVSGPGGLAAMSLGIPVYLHEONGIAGLTKMKLAKITKMOAFPGAPFAEVVGN
PYRTDVALPLPQOPLAGREGPVRYLVGDSGSGARLTNTPROVAKIGDSTVTHNS
GKSGSOSEDAIAEAGQPOHKYTERFIDMAAAYAMADVYVCSGALTVEIAGLPA
LSPFDHQRQOYWNALPLEKAGAKILIEPOLSDAVANLAWGSRRETLTMAERAR
ASIPDATERVANEVSRAVA"
92..96
/feature="put. ribosome binding site"
1182..1187
/feature="put. ribosome binding site"
1203..2678
/feature="murC"
1203..2678
/feature="murC"
/codon_start=1
/transl_table=11
/product="(UDP-N-acetylmuramate: L-alanine ligase) (AA
1..491)"
/protein_id="CA36868.1"
/db_xref="pid:842056"
/db_xref="gi:42056"
/db_xref="swiss-prot:P17952"
/translation="MNTQOLAKLRISIVPMRRVRHIFVIGAGGNGIAEVLNDEGY
QISGDLAPNPTQOIMNIGATIEYHNRPENTDASVYVSSAISDNDEIYAABAR
IVIRRAEMLAELMRERHGIAGTGKTTTAMSSIIYAEGALDPTFVNGGLARAG
VHARLHGRLYLADESDASFLHLOPMVAITYTTEADHMDYVGGDFEMLKQTFINEL
HNLPYGRAVMCVDDPYIRELLPRVGRQITTYGFSDDARVEDYQOITPQGHFTLLR
QKEPRKRVTLNAGRINALMAAANAVALTEBIDDAIILRALESQGTGRDFLGEF
PLEPANGSGITMLVDYDGHPTEDATIKARAGPDPNMLQPHFTTRDLYD
DEANVLTQVDTLMLEVYPAGAPPIGADSRSLCRITRGKIDPILVDPARVAEML
AVLNGNDLILVQAGANGICKIARSLAEIKLKQTPREEQHD"
2657..2661
/feature="murC"
/feature="put. ribosome binding site"
2671..2793
/codon_start=1
```

```
RESULT 13
LOCUS U32794 12085 bp DNA 09-SEP-1998
DEFINITION Haemophilus influenzae Rd section 109 of 163 of the complete
genome.
ACCESSION U32794 L42023
VERSION 91574694
KEYWORDS U32794.1 GI:1574694
SOURCE
ORGANISM
Haemophilus influenzae Rd.
Eubacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
REFERENCE 1 (bases 1 to 12085)
AUTHORS Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A.,
Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A.,
Merrick,J.M., McKenney,K., Sutton,G.G., Fitzhugh,W., Fields,C.A.,
Gocayne,J.D., Scott,J.D., Shirley,R., Lu,L.I., Glodek,A.,
Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E.,
Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M.,
Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L.,
Geoghagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M.,
Smith,H.O. and Venter,J.C.
TITLE Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd
JOURNAL Science 269 (5223), 496-512 (1995)
MEDLINE 95350630
REFERENCE 2 (bases 1 to 12085)
AUTHORS Tatusov,R.L., Muehlegan,A.R., Bork,P., Brown,N.P., Hayes,W.S.,
Borodovsky,M., Rudd,K.E. and Koonin,E.V.
TITLE Metabolism and evolution of Haemophilus influenzae deduced from a
whole-genome comparison with Escherichia coli
JOURNAL Curr. Biol. 6 (3), 279-291 (1996)
MEDLINE 96398784
REFERENCE 3 (bases 1 to 12085)
AUTHORS White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA
4 (bases 1 to 12085)
AUTHORS White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
TITLE Direct Submission
JOURNAL Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA
REMARK
Improved, and a number of frame shifts have been corrected. We
```



gratefully acknowledge the work of Tatusov et. al. We have incorporated their annotation into the /notes fields of the corresponding H. influenzae genes  
5 (bases 1 to 12085)  
REFERENCE White.O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D., Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.  
TITLE Direct Submission  
JOURNAL Submitted (28-MAY-1998) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
REMARK The whole genome was shifted by 588 nucleotides for a new start  
COMMENT On Oct 1, 1996 this sequence version replaced gi:1221895.  
FEATURES  
location/Qualifiers  
1..12085  
/organism="Haemophilus influenzae Rd"  
/db\_xref="taxon:71421"

## Source

gene

CDS

/transl\_table=11  
/product="UDP-N-acetylmutamate--alanine ligase (murC)"  
/protein\_id="AAC22794.1"  
/db\_xref="PID:g1574695"  
/db\_xref="GI:1574695"  
/translation="MKSHHEIRKTIIPMRVQOIHFIGIGAGMSGIAEILLNEGQ  
ISGSDIDAGVTOQLAQGAKITIGHAEHEHLEGASVYVSAIKDNEBELTSOKRI  
FVIOQADLAIEIMFRHGIAGVTHGKTITAMSMITQAKIDPTVNGVLVSACK  
NAHLAGSYLTAEDDESDFLHLQPMVSVTMMEDPDMDYEDGFEKAKTYVFLH  
NLPTGYALWCADDPVLMELVPRKQVITYGFEQADYRIEDYEOGFQGHYVICP  
NNEINLVNPGKHNLAATRALAVAKESGIANEALILEALADFGAGRRDGEET  
RPNCKVRLVDYGHHPTEVGVTKAARGMGDKRIVMFOPHRSRRDLFDDYVQL  
SOVALIMLDYANAGEAPTYGADSKSLCRSTRNGKVDPIVSDTSQGLVDQIID  
GDLIAGAGSVKISRGLESWKN"  
1598..2518  
/gene="H11140"  
1598..2518  
/note="similar to GB:D10483 SP:P07862 GB:K02668 PID:145724  
PID:146030 percent identity: 59.93; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="D-alanine--D-alanine ligase (ddlB)"  
/protein\_id="AAC22795.1"  
/db\_xref="PID:g1574696"  
/db\_xref="GI:1574696"  
/translation="MNLKOEKIAVLGTSAREVSINSGKAVLEALLKOGYNAPID  
PKRYNANLKKDGFNRANFNLHGRGEGDTMOGLLEOIGLPTGCGWASLLMDKRR  
TKMLKAFGLVAMKVTYRTRESELDPOAVAKLGIPLMKRPSLBSSGSIYKVA  
ELKSAYEALAKPENTILIEEMLAGDELYPVLDNOLVPAIRITVBESEPTDYKATYS  
DNITVFCPAGLTPEEQALSTLVRALDVGCRWSRIDWCDAKGNFRIVEVNTNG  
MTSHSLPFSATYGISFEOLVVKILESL"  
2518..3282  
/gene="H11141"  
2518..3282  
/note="similar to GB:D10483 SP:P06136 GB:K02668 GB:X02821  
PID:145725 percent identity: 40.62; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="cell division protein (ftsO)"  
/protein\_id="AAC22796.1"  
/db\_xref="PID:g1574697"  
/db\_xref="GI:1574697"  
/translation="MNLKRRKTPONIRRGOKPKYVPHIRAFVILGVFLGVYFW  
OSLEKMDKPRISAFVAGONTFTTADIKESLAKELGKFGQDAPVIOBOEARL  
WKAIVRKMPNRLSTWSTYOPAVWNOQVTLGIVFQULPSVRLTAKNPLYLG  
PDYSLKVIETWNIYINLKSNNIMAGINIDRGAVOVLDDNDIVLKLGRGWSKSL

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

ERFVITYPQIDVPEKNKIDYIDLRVTAGAAGWDR"  
3301..4578  
/gene="H11142"  
3301..4578  
/note="similar to GB:D10483 SP:P06137 GB:K02668 GB:M36531  
GB:X02821 percent identity: 52.84; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="cell division protein (ftsA)"  
/protein\_id="AAC22797.1"  
/db\_xref="PID:g1574698"  
/db\_xref="GI:1574698"  
/translation="MKGVETKTIIVGEGTSVWVVGEPFDPGVNVLGVGSPSK  
GIDRGSTDDAVGSIQRAIEAESADQKMSVLTALGHEHISQSLNEGFPVLES  
EVTOEIDSLHTASSIKLPEGLSLHVIPOEYAVDRQMKIKPGLQGVRLAQVHL  
IACHODMNNKRAKRCRGLOVDKVFESGFAATHSVLTEDKGLVCLIDFGATNV  
MYVTNGALRFSKVIYPAGNIVTNDIAHACTISPAEERIKVNVASAFPARLHDKKI  
EVASIGGARPSRLTKSDLSITSARTELGVYKDELKAKLEAKHIFELIAGV  
ITGGAGIIELEKCAASVHFQVRIASPLNITGLTVNRPQYSTVGLQVNVMSD  
DLAISGSDSEGFPEFSIMQGVKKIVKNSER"  
4662..5927  
/gene="H11143"  
4662..5927  
/note="similar to GB:D10483 SP:P06138 GB:X02821 PID:216509  
PID:40863 percent identity: 67.53; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="cell division protein (ftsZ)"  
/protein\_id="AAC22798.1"  
/db\_xref="PID:g1574699"  
/db\_xref="GI:1574699"  
/translation="MLPEYPEYDNFENESGALIKVVGSGGNAVHMMVNMVKQEM  
GGTFVSSSLTSEHGRIVYAVNTDQALRSQVOYQOIGETKRGAGANPVG  
RKAEDDODEIRKMLEGADVFIAGMGGTGGAAPVAKIKELGILTVAVTKF  
TFEGKKRMOEAIGIKDLQVYDSMTIIPNOIOIKVLPKNAKILDAFAANDLRMSY  
MGISDMITSPGLINVDPAVRVMSVQOGAMTGFSGVPGARAEAPARLAVRDL  
LEKIDISNMOGILVNTAGDLVFEERNIIGETIGSFASEEATVYVGTSLVPMSEI  
RVTIVATGGEINENPIQVROGLSTONIEGGRVNIIVELHRESVEVSRASEEY  
QRLPLDPIIDRLAFKKNFENPAREEN"  
5966..6883  
/gene="H11144"  
5966..6883  
/note="similar to GB:D10483 SP:P07652 PID:145848  
PID:216510 PID:40864 percent identity: 77.30; identified  
by sequence similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="UDP-3'-O-(3-hydroxymyristoyl) N-acetylglucosamine  
deacetylase (lpxC)"  
/protein\_id="AAC22799.1"  
/db\_xref="PID:g1574700"  
/db\_xref="GI:1574700"  
/translation="MTKORLKOSIVTGVGLHSGKVTLLTRPAMPNGVYVRRDL  
NPVAFPADNSVDRDTMLCTALINEDGVRSTVEHNLALDAGIDNITIEVADPFI  
IMDGASPFYLLDLAGIEQNNAKFIRIKOVRYVEDDKNAEFKPNYFRDLFDID  
FDHPAIGKDVNRNENMSAQAFFHQISRAETEFMDIDYLOSGLVIGGSLDNLAVL  
DDVRIENGLRKDELVRHKMDAIGDLYMAGYNIITGFKAVKSHGNNKLLRAVL  
ANQEAWEFVTEEDKAVPGGYAPGVLL"  
7010..8167  
/gene="H11145"  
7010..8167  
/note="similar to SP:002286 GB:M74134 GB:X60420 PID:148468  
PID:43344 percent identity: 55.32; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="chorismate mutase / prephenate dehydratase











- |   |   |   |  |
|---|---|---|--|
| REFERENCE<br>AUTHORS<br>TITLE                       | 6 (bases 1799 to 2187)<br>Haughn,G.W., Squitres,C.H., Defelice,M., Largo,C.T. and Calvo,J.M.<br>Unusual Organisation of the <i>ilvIH</i> promoter of <i>Escherichia coli</i><br>J. Bacteriol. 163 (1), 186-198 (1985)   | JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 18 (bases 16094 to 18886)<br>Ikeda,M., Machi,M., Jung,H.K., Ishino,F. and Matsushashi,M.<br>Nucleotide sequence involving murec and murec in the <i>mra</i> gene cluster<br>region of <i>Escherichia coli</i><br>Nucleic Acids Res. 18 (13), 4014 (1990)   |
| REFERENCE<br>AUTHORS<br>TITLE                       | 7 (bases 2101 to 4431)<br>Squires,C.H., Defelice,M., Devereux,J. and Calvo,J.M.<br>Molecular structure of <i>ilvIH</i> and its evolutionary relationship to<br><i>ilvIG</i> in <i>Escherichia coli</i> K12<br>Nucleic Acids Res. 11 (15), 5239-5313 (1983)  | JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 19 (bases 19464 to 21952)<br>Robinson,A.C., Kennan,D.J., Hatfull,G.F., Sullivan,N.F.,<br>Spiegelberg,R. and Donachie,W.D.<br>DNA sequence and transcriptional organization of essential cell<br>division genes <i>ftsQ</i> and <i>ftsA</i> of <i>Escherichia coli</i> : evidence for<br>overlapping transcriptional units<br>J. Bacteriol. 160 (2), 546-555 (1984) |
| JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 8 (bases 4274 to 6093)<br>Leclerc,G., Noel,G. and Drapeau,G.R.<br>Molecular cloning, nucleotide sequence, and expression of <i>shl</i> , a<br>new gene in the 2-minute region of the genetic map of <i>Escherichia</i><br><i>coli</i><br>J. Bacteriol. 172 (8), 4696-4700 (1990)  | JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 20 (bases 21464 to 23333)<br>Yi,Q.M. and Lutkenhaus,J.<br>The nucleotide sequence of the essential cell-division gene <i>ftsZ</i> of<br><i>Escherichia coli</i><br>Gene 36 (3), 241-247 (1985)   |
| JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 9 (bases 6088 to 7587)<br>Gomez,M.J., Flouret,B., van Heijenoort,J. and Ayala,J.A.<br>Nucleotide sequence of the regulatory region of the gene <i>pfbp</i> of<br><i>Escherichia coli</i><br>Nucleic Acids Res. 18 (9), 2813 (1990)  | JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 21 (bases 22964 to 25011)<br>Beall,B. and Lutkenhaus,J.<br>Sequence analysis, transcriptional organization, and insertional<br>mutagenesis of the <i>envA</i> gene of <i>Escherichia coli</i><br>J. Bacteriol. 169 (12), 5408-5415 (1987)  |
| JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 10 (bases 7316 to 10074)<br>Nakamura,M., Maruyama,I.N., Soma,M., Kato,J., Suzuki,H. and<br>Horoti,Y.<br>On the process of cellular division in <i>Escherichia coli</i> : nucleotide<br>sequence of the gene for penicillin-binding protein 3<br>Mol. Gen. Genet. 191 (1), 1-9 (1983)  | JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 22 (bases 23989 to 27799)<br>Schmidt,M.G., Rollo,E.E., Grodberg,J. and Oliver,D.B.<br>Nucleotide sequence of the <i>secA</i> gene and <i>secA(Ts)</i> mutations<br>preventing protein export in <i>Escherichia coli</i><br>J. Bacteriol. 170 (8), 3404-3414 (1988)   |
| JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 11 (bases 1 to 28277)<br>Tao,J.S. and Ishiguro,E.E.<br>Nucleotide sequence of the <i>mure</i> gene of <i>Escherichia coli</i><br>Can. J. Microbiol. 35 (11), 1051-1054 (1989)   | JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 23 (bases 27412 to 28277)<br>Akiyama,M., Horinuchi,T. and Sekiguchi,M.<br>Molecular cloning and nucleotide sequence of the <i>murT</i> mutator of<br><i>Escherichia coli</i> that causes A:T to C:G transversion<br>Mol. Gen. Genet. 206 (1), 9-16 (1987)  |
| JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 12 (bases 1 to 28277)<br>Michael,C., Parquet,C., Flouret,B., Blanton,D. and van Heijenoort,J.<br>Revised interpretation of the sequence containing the <i>mure</i> gene<br>encoding the UDP-N-acetylmuramyl-tripeptide synthetase of<br><i>Escherichia coli</i><br>Biochem. J. 269 (1), 277-278 (1990)  | JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 24 (bases 4274 to 6093)<br>Jahreis,K., Postma,P.W. and Lengele,J.W.<br>Nucleotide sequence of the <i>ilv H-frur</i> gene of <i>Escherichia coli</i> K-12<br>and <i>Salmonella typhimurium</i> LT2<br>Unpublished   |
| JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 13 (bases 11142 to 12634)<br>Parquet,C., Flouret,B., Mengin-Lecreux,D. and van Heijenoort,J.<br>Nucleotide sequence of the <i>mure</i> gene encoding the<br>UDP-murNAC-pentapeptide synthetase of <i>Escherichia coli</i><br>Nucleic Acids Res. 17 (13), 5379 (1989)  | JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 25 (bases 18619 to 19770)<br>Robinson,A.C., Kennan,D.J., Sweeney,J. and Donachie,W.D.<br>Further evidence for overlapping transcriptional units in an<br><i>Escherichia coli</i> cell envelope-cell division gene cluster: DNA<br>sequence and transcriptional organization of the <i>ddl ftsQ</i> region<br>J. Bacteriol. 167 (3), 809-817 (1986)                 |
| JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 14 (bases 12423 to 15030)<br>Ikeda,M., Machi,M., Ishino,F. and Matsushashi,M.<br>Nucleotide sequence involving <i>murec</i> and an open reading frame <i>ORF-X</i><br>spacing <i>murec</i> and <i>ftsW</i> in <i>Escherichia coli</i><br>Nucleic Acids Res. 18 (4), 1058 (1990)   | JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 26 (bases 20513 to 21772)<br>Robinson,A.C., Begg,K.J., Sweeney,J., Condie,A. and Donachie,W.D.<br>Mapping and characterization of mutants of the <i>Escherichia coli</i><br>cell division gene, <i>ftsA</i><br>Mol. Microbiol. 2 (5), 581-588 (1988)   |
| JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 15 (bases 13392 to 15020)<br>Mengin-Lecreux,D. and van Heijenoort,J.<br>Nucleotide sequence of the <i>mure</i> gene encoding the<br>UDP-murNAC- $\alpha$ -Ala-D-Glu synthetase of <i>Escherichia coli</i><br>Nucleic Acids Res. 18 (1), 183 (1990)  | JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | Erratum: [[published erratum appears in Mol Microbiol 1989<br>Feb;3(2):267]]<br>27 (bases 1 to 28277)<br>Wang,Q. and Calvo,J.M.<br>Ltp, a global regulatory protein of <i>Escherichia coli</i> , binds<br>cooperatively to multiple sites and activates transcription of<br><i>ilvIH</i><br>J. Mol. Biol. 229 (2), 306-318 (1993)                                  |
| JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 16 (bases 14743 to 16239)<br>Ikeda,M., Sato,T., Machi,M., Jung,H.K., Ishino,F., Kobayashi,Y. and<br>Matsushashi,M.<br>Structural similarity among <i>Escherichia coli</i> <i>ftsW</i> and <i>RodA</i> proteins<br>and <i>Bacillus subtilis</i> <i>SpoVE</i> protein, which function in cell<br>division, cell elongation, and spore formation, respectively<br>J. Bacteriol. 171 (11), 6375-6378 (1989) | JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 28 (bases 1 to 28277)<br>Wang,Q. and Calvo,J.M.<br>Ltp, a major regulatory protein in <i>Escherichia coli</i> , bends DNA and<br>can organize the assembly of a higher-order nucleoprotein structure<br>EMBO J. 12 (6), 2495-2501 (1993)   |
| JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE | 17 (bases 16094 to 17806)<br>Mengin-Lecreux,D., Texier,L. and van Heijenoort,J.<br>Nucleotide sequence of the cell-envelope <i>murec</i> gene of <i>Escherichia</i><br><i>coli</i><br>93285120  | JOURNAL<br>MEDLINE<br>REFERENCE<br>AUTHORS<br>TITLE |  |



COMMENT	This entry comprises a merged sequence of 28kb of which a portion is the submittor's original work.
FEATURES	location/Qualifiers
SOURCE	1. .2827

FEATURES	Location/Qualifiers
source	1. .28277

source 1. .28277

```
/organism="Escherichia coli"  
/strain="K12 W3110"
```

```
/sclname="K12 W3110"
/db_xref="taxon:562"
```

/clone="PLC26-6, PLC4-14, PLC

```

/mmap="1.77 to 2.38 mins"
complement(1 359)

```

complement(<l.:269)  
/note="leu mRNA"  
mRNA

/evidence=experimental

```
mat_peptide      complément(<1. .63)
```

...  
Note: remainder of annotations omitted.



\*\*\*\*\*  
WISREH (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_PP protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Nov 24 00:43:15 1999; MasPar time 0.52 Seconds  
326.236 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-103-287-2  
Description: (1-437) From US09103287.pep  
Perfect Score: 3121  
Sequence: 1 MTHYFVGKSGMSLAQI.....GDIQKLNAYLDKLGKNAF 437

Scoring table: PAM 150  
Gap 11

Searched: 1479 seqs, 387401 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: HIV-AA8  
1:ALL

Statistics: Mean 46.900; Variance 101.768; scale 0.461

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	88	2.8	1006	1 POLSMN	HIVMN, POL POLYPROTEI	6.31e+01
2	88	2.5	489	1 GAGSVI57	HIVVI57, GAG POLYPROT	5.92e+00
3	74	2.4	112	1 ENVSFLPD9	HIVFLPD9, ENV POLYPR	1.34e+01
4	76	2.4	115	1 ENVSFLPD9	HIVFLPD9, ENV POLYPR	8.96e+00
5	74	2.4	116	1 ENVSFLPD1	HIVFLPD1, ENV POLYPR	1.34e+01
6	76	2.4	340	1 NEFSJB03	HIVJB03, NEF	8.96e+00
7	76	2.4	856	1 ENVSFLVAM1	FIIVAM19, ENV POLYPROT	8.96e+00
8	75	2.4	1003	1 POLSOXI	HIVOXI, POL POLYPROTE	1.10e+01
9	75	2.4	1003	1 POLSCAM1	HIVSCAM1, POL POLYPROT	1.10e+01
10	75	2.4	1003	1 POLSSF2	HIVSSF2, POL POLYPROTE	1.10e+01
11	73	2.3	148	1 ENVSFLPD3	HIVFLPD3, ENV POLYPRO	1.63e+01
12	72	2.3	148	1 POLS26	HIV26, POL POLYPROTEI	1.98e+01
13	71	2.3	264	1 GAGSVI354	HIVVI354, GAG POLYPRO	2.39e+01
14	71	2.3	385	1 GAGSVI32	HIVVI32, GAG POLYPROT	1.63e+01
15	73	2.3	486	1 GAGSVI415	HIVVI415, GAG POLYPRO	1.63e+01
16	71	2.3	489	1 GAGSVI245	HIVVI245, GAG POLYPRO	2.39e+01
17	73	2.3	490	1 GAGSLBV23	HIVLBV23, GAG POLYPRO	1.63e+01
18	72	2.3	853	1 ENVELOPESA	HIVADA, ENVELOPE POLY	1.98e+01
19	72	2.3	1002	1 POLSNDK	HIVNDK, POL POLYPROTE	1.98e+01
20	72	2.3	1002	1 POLS2226	HIV2226, POL POLYPROTE	1.98e+01
21	72	2.3	1002	1 POLSHAN	HIVSHAN, POL POLYPROTE	1.98e+01
22	72	2.3	1002	1 POLSMAL	HIVMAL, POL POLYPROTE	1.98e+01
23	71	2.3	1002	1 POLSU455	HIVU455, POL POLYPROT	2.39e+01

24	73	2.3	1003	1 POLSNV5CG	HIVNV5CG, POL POLYPRO	1.63e+01
25	73	2.3	1003	1 POLSNL43	HIVNL43, POL POLYPROT	1.63e+01
26	72	2.3	1003	1 POLSBCG3C	HIVBCG3C, POL POLYPR	1.98e+01
27	72	2.3	1003	1 POLSHXB2	HIVHXB2, POL POLYPROT	1.98e+01
28	72	2.3	1003	1 POLSJREL	HIVJREL, POL POLYPROT	1.98e+01
29	72	2.3	1003	1 POLSHXB2R	HIVHXB2R, POL POLYPRO	1.98e+01
30	72	2.3	1003	1 POLSYU2	HIVYU2, POL POLYPROTE	1.98e+01
31	72	2.3	1003	1 POLSD31	HIVD31, POL POLYPROTE	1.98e+01
32	72	2.3	1007	1 POLVJRCSE	HIVVJRCSE, POL POLYPRO	1.98e+01
33	72	2.3	1015	1 POLSLAI	HIVSLAI, POL POLYPROTE	1.98e+01
34	72	2.3	1015	1 POLSBRU	HIVBRU, POL POLYPROTE	1.98e+01
35	73	2.3	1109	1 POLSCAEVCG	CAEVCG, POL POLYPROTE	1.63e+01
36	69	2.2	99	1 POLSJH31	HIVJH31, POL POLYPROT	3.47e+01
37	68	2.2	103	1 GAGSFL10S1	HIVFL10S1, GAG	4.15e+01
38	70	2.2	242	1 GAGSIC144	HIVIC144, GAG	2.89e+01
39	70	2.2	304	1 GAGSIC159	HIVIC159, GAG POLYPROT	2.89e+01
40	69	2.2	486	1 GAGSDJ258	HIVDJ258, GAG POLYPRO	3.47e+01
41	69	2.2	488	1 GAGSVI310	HIVVI310, GAG POLYPRO	3.47e+01
42	70	2.2	492	1 GAGSK124	HIVK124, GAG POLYPROT	2.89e+01
43	68	2.2	856	1 ENVELOPESF	FIIVT113, ENVELOPE PO	4.15e+01
44	69	2.2	1002	1 POLSRF	HIVRF, POL POLYPROTEI	3.47e+01
45	70	2.2	1045	1 POLSVLVC	VLVNC, POL POLYPROTEI	2.89e+01

## ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	1006	AA
1	1 POLSMN				
AC	M17449				
DE	01-MAR-1989				
DT	HIVMN, POL POLYPROTEIN				
FT	PEPT				
CC	-1-5'-N 3'-Y 51				
CC	TRANSLATED USING PHASE 1				
CC	SEQUENCE 1006 AA: 113750				
MM:					

Query Match 2.8%; Score 88; DB 1; Length 1006;  
Best Local Similarity 18.3%; Pred. No. 6.31e+01;  
Matches 20; Conservative 35; Mismatches 47; Indels 7; Gaps 7;

DB	34	grdnmsiseageeagddtqgp-v-sfsfpqtlwprlvltkigqg-1kealldtqaddt 90
QY	11	GGMSLSLAQIMHDLHEVCGSDIENVFVETALRNKKIKLDPFGANNIKEDWVVIQGNAF 70
DB	91	vlgemmlprwkpkm-igsgigfi-kvr-qvqgdlci-gicphkxagtvy 135
QY	71	ASSHEIYNAHOLKLDVVSYNDFLQIIDIOTYSVAVTGAGHKTSTGTL 119

RESULT	ID	PRELIMINARY	PRT	489	AA
2	1 GAGSVI57				
AC	L11794				
DT	20-APR-1993				
DE	HIVVI57, GAG POLYPROTEIN				
FT	PEPT				
CC	-1-5'-Y 3'-N 1 >489				
CC	TRANSLATED USING PHASE 1				
CC	SEQUENCE 489 AA: 54331				
MM:					

Query Match 2.5%; Score 78; DB 1; Length 489;  
Best Local Similarity 22.0%; Pred. No. 5.92e+00;  
Matches 9; Conservative 18; Mismatches 11; Indels 3; Gaps 3;

RESULT	ID	PRELIMINARY	PRT	112	AA
3	1 ENVSFLPD9				
AC	M90885				
DT	03-FEB-1992				
DE	HIVFLPD9, ENV POLYPROTEIN				



DI	01-MAR-1989			
DE	HIVSF2, POL	POLYPROTEIN		
FT	PEPT	51	100303	1 1003 POL







**THIS PAGE BLANK (USPTO)**



(W.T.)

Distribution rights by Oxford Molecular Ltd

which has been backtranslated into n.a. using IUPAC symbols

Tabular output not generated.

Comp:

Match STD :

Listing first 45

Database: HIV-NAS

### Statistics:

and is derived by analysis of the total score distribution.

## SUMMARIES

19	116	2.1	1459	3	HIVDJ258	Human immunodeficiency	5.66e+00
----	-----	-----	------	---	----------	------------------------	----------

c	45	111	2.0	7817	2	HIV2D205	Human immunodeficienc	1.40e+01
---	----	-----	-----	------	---	----------	-----------------------	----------

## ALIGNMENTS

BAST COUNT	344 a	281 c	369 g	286 t
------------	-------	-------	-------	-------



ORIGIN	gag cds start	2.6%;	Score 141;	DB 3;	Length 1480;
Query Match					
Best Local Similarity	43.1%;		Pred. No. 2.65e-02;		
Matches	31;	Conservative	14;	Mismatches 27;	Indels 0;
Gaps	0;				
Db	269	aaagatagatgctaaagacacccaagaagctttagagaagtgtagagaagacaaca	328		
Y	833	ARAAATYTGAGTGNACNCAATATHAARGAGCNYTNGARACATTGTGGGNGGTNAARGM	892		
Db	329	aaagtaagga	340		
Y	893	GNTTYAAGARA	904		
RESULT	2				
LOCUS	HIVM251	303 bp ss-RNA	VR	09-SEP-1993	
DEFINITION	Human immunodeficiency virus type 1, clone 2, ADC sample M, partial gag cds, p17 region.				
ACCESSION	L21492				
SOURCE	1 of 4				
ORGANISM	Human immunodeficiency virus type 1 (HIV-1), clone 2, ADC sample M. Human immunodeficiency virus type 1				
REFERENCE	1 (bases 1 to 303)				
AUTHORS	Zhu,T., Mo,H., Wang,N., Nam,D.S., Cao,Y., Koup,R.A. and Ho,D.D.				
TITLE	Genotypic and phenotypic characterization of HIV-1 in patients with primary infection				
JOURNAL	Science 261, 1179-1181 (1993)				
COMMENT	full staff review				
FEATURES	This sample has been taken from patient M, one of the seven patients (A,M,L,F,C,V,R), in the study(1). All patients had acute, self-limited symptomatic illness with measurable viremia followed by seroconversion. Patient M was a chronically infected male who subsequently transmitted the virus to patient F. The authors report that the sequences in this study are found to be uniformly macrophage-tropic and non-syncytium-inducing. The sequence for sample M clone 2, along with other patient M clones 1, 3-9, 11, and 12, is shown in alignment following the printed text entry. These gag sequences cluster with HIV-1 B subtype sequences. See also L21224-L21591 and L24161, L24162.				
CDS	location/Qualifiers				
	<1..>303				
	/product="myristylated gag protein p17"				
	/gene="gag"				
	/codon_start=1				
	<1..>303				
	/organism="Human immunodeficiency virus type 1"				
	/sequenced_mol="RNA"				
source					
BASE COUNT	125 a 61 c 62 g 55 t				
ORIGIN	190 bp downstream from beginning of gag cds				
Query Match	2.4%;	Score 131;	DB 3;	Length 303;	
Best Local Similarity	41.7%;	Pred. No. 2.65e-01;			
Matches	30;	Conservative	14;	Mismatches 28;	Indels 0;
Gaps	0;				
Db	80	aaagatagatgctaaagacacccaagaagctttagagaagatagagaagacaaca	139		
Y	833	ARAAATYTGAGTGNACNCAATATHAARGAGCNYTNGARACATTGTGGGNGGTNAARGM	892		
Db	140	aaagtaagga	151		
Y	893	GNTTYAAGARA	904		
RESULT	3				
LOCUS	HIVF1051	309 bp ss-RNA	VR	09-SEP-1993	
DEFINITION	Human immunodeficiency virus type 1, clone 10, ADC sample F, partial gag cds, p17 region.				
ACCESSION	L21375				
SEGMENT	1 of 4				
SOURCE	Human immunodeficiency virus type 1 (HIV-1), clone 10, ADC sample F.				

```

ORGANISM
Human immunodeficiency virus type 1
Virus; ss-RNA enveloped viruses; Positive strand RNA virus;
Retroviridae; Lentivirinae.

REFERENCE
1 (bases 1 to 309)
AUTHORS
Zhu,T., Mo,H., Wang,N., Nam,D.S., Cao,Y., Koup,R.A. and Ho,D.D.
TITLE
Genotypic and phenotypic characterization of HIV-1 in patients
with primary infection
JOURNAL
Science 261, 1179-1181 (1993)
STANDARD
full staff_review
COMMENT
This sample has been taken from patient F, one of the seven
patients (A,M,L,F,C,V,R), in the study(1). All patients had acute,
self-limited symptomatic illness with measurable viremia followed
by seroconversion. Patient F was infected by patient M via sexual
transmission. The authors report that the sequences in this study
are found to be uniformly macrophage-tropic and nonsyncytium-
inducing. The sequence for sample F clone 10, along with other
patient F clones 1-9, 11, and 12, is shown in alignment following the
printed text entry. These gag sequences cluster with HIV-1 B subtype
sequences. See also L21224-L21591 and L24161, L24162.

FEATURES
CDS
location/Qualifiers
<1..>309
/product="myristylated gag protein p17"
/gene="gag"
/codon_start=1
/translation="LOPSLQTSSELSLYMTIAVLYCVOHKIDVKDTKEALEKEE
ONSKRKAQDQATRAADDTANSSVGSNRYIVNIGQGMVHQISRTILNANVYEEENA
F"
source
<1..>309
/organism="Human immunodeficiency virus type 1"
/sequenced_mol="RNA"
BASE COUNT
128 a 64 c 62 g 55 t
ORIGIN
190 bp downstream from beginning of gag cds

Query Match
2.4%; Score 131; DB 3; Length 309;
Best Local Similarity 41.7%; Pred. No.2,65e-01;
Matches 30; Conservative 14; Mismatches 28; Indels 0; Gaps 0;

Db
140 aaagtaagaana 151
::: || |
Gy 893 GNTTYAAGARA 904

RESULT 4
LOCUS HIVPH136 1474 bp ss-RNA VRL 20-APR-1993
DEFINITION Human immunodeficiency virus type 1, Philippines isolate PH136, gag
region.
ACCESSION L11780
SOURCE Human immunodeficiency virus type 1 (HIV-1), Philippines isolate
PH136.
ORGANISM Human immunodeficiency virus type 1
Virus; ss-RNA enveloped viruses; Positive strand RNA virus;
Retroviridae; Lentivirinae.
REFERENCE
1 (bases 1 to 1474)
AUTHORS
Lounagile,J.J., McCutchan,F., Brennan,T., Peeters,M., Brennan,T.,
Ganders-Buell,E., Eddy,G., van der Groen,G., Fransen,K.,
Gershy-Damet,M., Deleys,R. and Burke,D.
TITLE
Phylogenetic analysis of gag genes from seventy international HIV-1
isolates provides evidence for multiple genotypes
JOURNAL
AIDS 7, 769-780 (1993)
STANDARD
full staff_review
COMMENT
Kindly provided prior to publication by Henry M. Jackson Foundation
Research Laboratory, 1500 East Gude Drive, Rockville, MD 20850.
The PH136 gag sequence clusters with HIV-1 B subtype sequences.

FEATURES
CDS
location/Qualifiers
1..1474
/product="gag protein"
/gene="gag"
/codon_start=1
/translation="WGARASVLSGGELDRWRERILRPGCKKKYKLKHITWASELEERF"

```







exon	8380..>8470	/number=2	/product="tat protein"
exon	8380..>8654	/number=2	/product="rev protein"
CDS	8798..9430	/product="nef protein"	/codon_start=1
LTR	9099..9553	/translation="MGGRKSRSLGGSVAVERMORAEPRAPAEVGAVSRLERN GATTSNTAATNADCAWLEAQEEVEGFPAPQVPLRMTYKALDLSHLKEGGLDE GLIISORODILDEIMYITGCFPDMQNYTPGPRIPLFGWCFKLPVPEQVEEBA NKRNASALHPMSQGHMDPKEVKLMMKFDSRLAFHMAARKHPEFKDC" 9099..9553	
BASE COUNT	3405 a 1709 c 2336 g 2141 t		
ORIGIN			
Query Match	2.4%; Score 131; DB 3; Length 9591;		
Best Local Similarity	41.7%; Pred. No. 2,65e+01;		
Matches	30; Conservative 14; Mismatches 28; Indels 0; Gaps 0;		
Db	1060 aaaaagatagytlaaagaccaccaaagagcccttagagaaatagaggaagcaaaa 1119		
Oy	833 ARAARYTNGAYTNCNMAVATHAARGACGNYTGARACNTTYGNGNGNGNARMGMN 892		
Db	1120 aaagtaagaaaa 1131		
Oy	893 GNTTYAAGARA 904		
RESULT	6		
LOCUS	HIVC159 1459 bp ss-RNA	VRL	20-APR-1993
DEFINITION	Human immunodeficiency virus type 1, isolate C159 from Abidjan, Ivory Coast, gag region.		
ACCESSION	U11759		
SOURCE	Human immunodeficiency virus type 1 (HIV-1), Abidjan, Ivory Coast isolate C159 (cocultivation on PBMCs).		
ORGANISM	Human immunodeficiency virus type 1		
REFERENCE	Viridae; ss-RNA enveloped viruses; Positive strand RNA virus; Retroviridae; Lentivirinae.		
AUTHORS	1 (bases 1 to 1459)		
TITLE	Louwagie,J.J., McCutchan,F., Brennan,T., Peeters,M., Brennan,T., Sanders-Buell,E., Eddy,G., van der Groen,G., Franssen,K., Gersh-Damet,M., Deleys,R. and Burke,D.		
JOURNAL	Phylogenetic analysis of gag genes from seventy international HIV-1 isolates provides evidence for multiple genotypes		
STANDARD	AIDS 7, 769-780 (1993)		
COMMENT	Kindly provided prior to publication by Henry M. Jackson Foundation Research Laboratory, 1500 East Gude Drive, Rockville, MD 20850. This sequence contains a premature stop codon at position 913. The c159 gag sequence clusters with HIV-1 A subtype sequences.		
FEATURES	Location/Qualifiers		
CDS	1..912		
	/note="premature termination at 913"		
	/product="gag protein"		
	/codon_start=1		
	/translation="MGARASVLSGGKLDAMEKIRLPGRGKKYRIKHLVNASRELRF ALNGLLETRADGCOQLMEQLOASLRITSEELKSFNAIYLYKCYHORIDIKDKREALD KLEEVNKRKQTDQAAATGSGQNPPIVONOGQMTHTQPSBPRIINAWKVIIEKA FSPVIMKPIALSGATPODLMNMLITVGHQAMOMLKDPTINEEAMDRVHPVAG		

[illegible]



RESULT	LOCUS	DEFINITION	ACCESSION	ORGANISM	REFERENCE	TITLE	JOURNAL	STANDARD	COMMENT
9	HIV20667K	766 bp ds-DNA	VBL	23-AUG-1994					
		Human immunodeficiency virus type 2 (HIV-2) p16/p28 gag sequence, sample 60667K, clone 1, partial cds.							
		133078							
		Human immunodeficiency virus type 2 PCR-amplified DNA recovered from a 37 year old heterosexual female originally from Sierra Leone who showed HIV dual seroreactivity.							
		Human immunodeficiency virus type 2							
		Viralae; ss-RNA enveloped viruses; Positive strand RNA virus; Retroviridae; Lentivirinae.							
		1 (bases 1 to 766)							
		Gao,F., Yue,L., Robertson,D.L., Hill,S.C., Hui,H., Biggar,R.J., Neegaaye,A.E., Whelan,T.M., Ho,D.D., Shaw,G.M., Sharp,P.M. and Hahn,B.H.							
		Genetic diversity of human immunodeficiency virus type 2: Evidence for five distinct sequence subtypes with differences in virus biology							
		J. Virol. 68, 7433-7447 (1994)							
		Full staff review							
		This sequence was obtained as part of a study designed to determine the in vivo extent of HIV-2 genetic and biological variation. Up to now, HIV-2 sequences have been derived almost solely from cultured isolates. Since cultivation may select against viruses which can not grow in a particular cell type, Gao et al. argue that a greater variation may be elucidated by sequencing virus obtained from uncultured PBMCs. With this aim in mind, peripheral blood samples were collected from 12 HIV-2 seropositive patients (2238, 6041K, 60667K, 7312A, 7810A, 7924A, FA, F0784, FT, JA, ON, and PA), from six different West African countries, living in both urban and rural areas. Uncultured PBMC DNA was PCR amplified, and the recombinant clones were manually sequenced. Another portion of the cells was used for virus isolation.							
		Patient 60667K was a heterosexual 37-year old female from Sierra Leone, who was living in the Washington, D.C. area. Originally she lived in an urban area in West Africa before traveling extensively and coming to the United States. After seeking medical care for a STD, patient 60667K was identified as HIV-2 positive. She had reduced CD4 counts but was otherwise asymptomatic. Her two teenage children were seronegative. Western blot analysis of patient 60667K's peripheral blood cells showed HIV dual reactivity. Virus isolation was successfully performed at the University of Alabama a Birmingham. Sample 60667K clusters with the subtype B3 viruses.							
		NBCI gi: 532133							
		Location/Qualifiers							
		1..766							
		/organism="Human immunodeficiency virus type 2"							
		/proviral							
		/sequenced_mol="DNA"							
		/tissue_type="PBMC"							
		<1..>766							
		/partial							
		/gene="gag"							
		/note="NCBI gi: 532134"							
		/codon_start=2							
		/translation="VNELERFGAENILGSECKHILAVLEPLVPTGSENLSKLENTVCVYICILHAGEKVKDTEAKRIKRIORHIVGTETEMPATSRPTASSGNGYVOOIAAGNYSHLPSPRTNIAWVKIVEEKKFGENVPGVQALSEGTPDINOMLNCVGDHDAIETIIIRIIRIEEAEMDQHPSPGPMAGOLRERGSIDINGTSTVEEQIOMCRADNPVVGNTIRRMQIGLOKCYRMATNPVITILDKQPKKEPFON"							
		BASE COUNT							
		275 a 160 c 187 g 144 t							
		ORIGIN							
		111 bp downstream from the start of gag							
		Query Match							
		2.2% Score 119; DB 3; Length 766;							
		Best Local Similarity 46.8% Pred. No. 3.1e+00;							
		Matches 22; Conservative 11; Mismatches 14; Indels 0; Gaps 0;							
		Db 410 gaagagatgcaccccatgatataatcagatgttaattgtgag 456							
		:::::							
		Cp 1185 RAASWNGCNCNCNDATYTRTCDATNARTCTGTGATGNTMARNG 1139							



LOCUS	HIVV1557	1468 bp ss-RNA	VRL	20-APR-1993
DEFINITION	Human immunodeficiency virus type 1, zairean isolate V1557, gag region.			
ACCESSION	L11793			
SOURCE	Human immunodeficiency virus type 1 (HIV-1), Zairean isolate V1557.			
ORGANISM	Human immunodeficiency virus type 1			
REFERENCE	Vittidee; ss-RNA enveloped viruses; Positive strand RNA virus; Retroviridae; Lentivirinae.			
AUTHORS	1 (bases 1 to 1468) Louwagie,J.J., McCutchan,F., Brennan,T., Peeters,M., Brennan,T., Sanders-Buell,E., Eddy,G., van der Groen,G., Franssen,K., Gershby-Damet,M., Deleys,R. and Burke,D.			
TITLE	Phylogenetic analysis of gag genes from seventy international HIV-1 isolates provides evidence for multiple genotypes AIDS 7, 769-780 (1993)			
JOURNAL				
STANDARD				
COMMENT	Kindly provided prior to publication by Henry M. Jackson Foundation Research Laboratory, 1500 East Gude Drive, Rockville, MD 20850. The authors group the outlying V1557 gag sequence with V1525 as subtype G.			
FEATURES	Location/Qualifiers			
CDS	1..1468 /product="gag protein" /gene="gag" /codon_start=1 /translation="MGARASYLSGGKLDAMEKILRPQGGKKRYIKLHWASRELERRALNLDLTDAAGCQLLEQLPALKTGTEELSGFNLATLYCHORIDVDYKEALEKLAKAESONRIQOATGDGNKGNIISNTYITVOAGOPHQAISRPLTAAWVVEEKAASPEVTIPMFSALESBGAIPQDNLKMIINYGOAOAMOMKDIINEEAEMDVHPVHAGCIPIPGMOEPREGSDIAGTSILODIAMTSNPPIPDIKKIRIIIGLKITVMKSPSIILDIRGPREPRDYIDRFKALRAAQAOEVKGMWTDLIVONANPDKTILLRALQGASIEEMNIAACGVGGPGHKAVLAFAWSOVNTVAWMOKNFNGQRITICKFRNCKEHLARNCRAPRGKCMKCCKGHOLKDCTEROANTGLKIMSPNGRGPENFLQSREPTAPPAPESEFGEGEMTPPPQERKDEASPLTSKSX" 1..1468 /organism="Human immunodeficiency virus type 1" /isolate="V1557" /cell_type="Tymphocyte" /proviral /sequenced_mol="DNA" /tissue_type="blood"			
BASE COUNT	538 a 277 c 363 g 290 t			
ORIGIN	gag cds start			
Query Match	2.2%; Score 122; DB 3; Length 1468;			
Best Local Similarity	46.6%; Pred. No. 1.76e+00;			
Matches	Conservative 13; Mismatches 17; Indels 1; Gaps 1;			
Db	255 ctatcgctacatacaagaat-agattgtcaaaagcaccaaggagccttagagaagt 311    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::   QY 818 CNAITHWSTAYTTNGARARYTNAGYGTMACMAAATHTAARGARCNTNGARACNTT 875			
RESULT 11				
LOCUS	HIVPH153	1492 bp ss-RNA	VRL	20-APR-1993
DEFINITION	Human immunodeficiency virus type 1, Philippines isolate PH153, gag region.			
ACCESSION	L11781			
SOURCE	Human immunodeficiency virus type 1 (HIV-1), Philippines isolate PH153.			
ORGANISM	Human immunodeficiency virus type 1			
REFERENCE	Vittidee; ss-RNA enveloped viruses; Positive strand RNA virus; Retroviridae; Lentivirinae.			
AUTHORS	1 (bases 1 to 1492) Lowwagie,J.J., McCutchan,F., Brennan,T., Peeters,M., Brennan,T., Sanders-Buell,E., Eddy,G., van der Groen,G., Franssen,K., Gershby-Damet,M., Deleys,R. and Burke,D.			
TITLE	Phylogenetic analysis of gag genes from seventy international HIV-1 isolates provides evidence for multiple genotypes AIDS 7, 769-780 (1993)			
JOURNAL				
STANDARD				
COMMENT	Kindly provided prior to publication by Henry M. Jackson Foundation Research Laboratory, 1500 East Gude Drive, Rockville, MD 20850.			

The PH153 gag sequence clusters with HIV-1 B subtype sequences.

FEATURES  
CDS  
1..1474  
/product="gag protein"  
/gene="gag"  
/codon\_start=1  
/translation="MGARASVLSGKRLDRREKTRLRAPGCKKKQKLHIYASRELF  
AVNGELIETSEGCROIUGLHPALDGTETELRLNLTVALTVHONIEYKDKLE  
KIEEQKSNKSKKQAQAADGSCSVSONPVIYONIGOVHVAISPRILNAAWKVE  
EKAFSPVIMFSAKLSLEGATPODINTMLNTVNGCHQAMOMLKETINEEAEMRLHPV  
HAGIAGGOKREPPGSDIAGTSTLQIGAMNNPPIPGVETIKMEILGLKIVM  
YSPSTLIDROGKREPPRODVDRFYKTLRAEOASODVKMMETLLVONANPCKTIL  
KALGRATLLEMTACGVCGPGHKARVLAELASQATNSATMLQGNFRNRTKYKC  
FNCKEGBHIANCKAPRKKCGMCGRGHOMKDCTERGQATVFLKIMPSHGRPNFLQ  
SRPEPTAPPEESFRRETTTPPOKEPDIKELPSASLKLX"  
1..1474  
/organism="Human immunodeficiency virus type 1"  
/proviral  
/sequenced\_mol="DNA"  
1..1474  
/organism="Human immunodeficiency virus type 1"  
/isolate="PH153"  
/cell\_type="lymphocyte"  
/proviral  
/sequenced\_mol="DNA"  
/tissue\_type="blood"  
BASE COUNT 557 a 295 c 357 g 283 t  
ORIGIN gag cds start

Query Match 2.2% Score 121; DB 3; Length 1492;  
Best Local Similarity 41.7%; Pred. No.2.15e+00; Indels 0; Gaps 0;  
Matches 30; Conservative 13; Mismatches 29; Indels 0; Gaps 0;

Db 269 aaacatagaggtaaagacacccaagaagactttagaagaatagaggaagacaaaca 328  
|:|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| :  
QY 833 ARAAATYTGATGATNCAACNAATATHARGARGCNGTNGARACHTTGTGGNGGCTNARAGNM 892

Db 329 aaagtacaacaaa 340  
|:|:| |:| :  
QY 893 GNTVYAAGARA 904

RESULT 12  
LOCUS HIVTB1321 1492 bp ss-RNA VRL 11-JUL-1992  
DEFINITION Human immunodeficiency virus type 1, sample BK132 from Bangkok,  
Thailand, gag region.  
ACCESSION J03696  
SEGMENT 1 of 2  
SOURCE Human immunodeficiency virus type 1 (HIV-1), Bangkok, Thailand  
sample BK132, PCR amplified and cloned after coculture on PBMCs.  
ORGANISM Human immunodeficiency virus type 1  
Virusidae: ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae: Lentivirinae.  
1 (bases 1 to 1492)  
REFERENCE 1 (bases 1 to 1492)  
AUTHORS McCutchan, F.E., Hegerich, P.A., Brennan, T.P., Phanuphak, P.,  
Sungtharaj, P., Jungsuddee, A., Berman, P.W., Gray, A.M., Fowler, A.K. and  
Burke, D.S.  
TITLE Genetic variants of HIV-1 in Thailand  
JOURNAL AIDS Res. Hum. Retroviruses (1992) In press  
STANDARD full staff\_entry  
COMMENT kindly provided prior to publication by Henry M. Jackson Foundation  
and the Walter Reed Army Institute of Research, Rockville, MD 20850.  
FEATURES The 132 sequence clusters with HIV-1 B subtype gag sequences.  
pept from to/cspn description  
BASE COUNT 559 a 286 c 362 g 285 t  
ORIGIN

Query Match 2.2% Score 121; DB 3; Length 1492;  
Best Local Similarity 41.7%; Pred. No.2.15e+00; Indels 0; Gaps 0;  
Matches 30; Conservative 13; Mismatches 29; Indels 0; Gaps 0;

Db 269 aaacatagaggtaaagacacccaagaagactttagaagaatagaggaagacaaaca 328



QY	833	ARAAATYINGAGTACCAAAATATTAAGACRCNTTNGARACNTTGGNGNGTNAARMGNM	892
Db	329	aaagtaagaanaa 340	
QY	893	GNTTYAAYGARA 904	
RESULT	13		
LOCUS	HIVBCSG3C	9168 bp ss-RNA	VRL 28-JUL-1993
DEFINITION	Human immunodeficiency virus type 1, isolate BC, molecular clone SG3, complete genome.		
ACCESSION	L02317		
SOURCE	Human immunodeficiency virus type 1 (HIV-1), isolate BC, molecular clone SG3.		
ORGANISM	Human immunodeficiency virus type 1		
REFERENCE	Viridae: ss-RNA enveloped viruses; Positive strand RNA virus; Retroviridae: Lentivirinae.		
AUTHORS	1 (bases 1 to 9168) Ghosh,S.K., Fultz,P.N., Keddie,E., Sharp,P.M., Saag,M.S., Hahn,B.H. and Shaw,G.M.		
TITLE	A molecular clone of HIV-1 tropic and cytopathic for human and chimpanzee lymphocytes		
JOURNAL	Virology 194, 858-864 (1993)		
STANDARD	full statf-review		
COMMENT	kindly provided in computer-readable form by Dr. Beatrice Hahn, CCR, University of Alabama at Birmingham, Birmingham AL. This represents one of the most cytopathic variants that the authors of [1] have encountered. The BC virus is cytopathic in both human and chimp lymphocytes. This sequence clusters with HIV-1 subtype B sequences. SG3 is T-cell tropic as opposed to monocytropic [1].		
FEATURES	Location/Qualifiers		
misc-feature	1..9168		
	/note="entire sequence was obtained from replication competent plasmid clone pSG3.1"		
source	1..9168		
	/organism="Human immunodeficiency virus type 1"		
	/isolate="BC"		
	/strain="SG3"		
	/cell_line="Immortalized T-cell"		
	/cell_type="lymphoid"		
	/proviral		
	/sequenced_mol="DNA"		
BASE COUNT	3291 a 1633 c 2225 g 2019 t		
ORIGIN	beginning of R region of the 5' LTR		
Query Match	2.2%; Score 121; DB 3; Length 9168;		
Best Local Similarity	40.3%; Pred. No.2.15e+00;		
Matches	29; Conservative 14; Mismatches 29; Indels 0; Gaps 0;		
Db	605	aatgatagatgtaagaacacacagaagacttagagaagaatagaagaagaacaaaca 664	
QY	833	ARAAATYINGAGTACCAAAATATTAAGACRCNTTNGARACNTTGGNGNGTNAARMGNM 892	
Db	665	acagtaagaanaa 676	
QY	893	GNTTYAAYGARA 904	
RESULT	14		
LOCUS	HIV2D194	9472 bp ss-RNA	UNA 12-MAR-1991
DEFINITION	Human immunodeficiency virus type 2 (HIV-2), isolate D194 from a Gambian person; complete proviral genome.		
ACCESSION	J04542	X52223	
SOURCE	Human immunodeficiency virus type 2, isolate D194, clone 10; derived from circular proviral DNA; biologically active clone.		
REFERENCE	1 (bases 1 to 9399)		
AUTHORS	Blesert,L., Kreutz,R., Immelmann,A., Henco,K., Weichsner,C., Andresen,R., Gelderblom,H. and Rubsamen-Waigmann,H.		
TITLE	Molecular cloning of two West African human immunodeficiency virus type 2 isolates that replicate well in macrophages: a Gambian		

JOURNAL STANDARD REFERENCE AUTHORS	TITLE	JOURNAL STANDARD COMMENT
isolate, from a patient with neurological acquired immunodeficiency syndrome, and a highly divergent Ghanaian isolate Proc. Natl. Acad. Sci. U.S.A. 86, 2383-2387 (1989)	unpublished (1991) full staff review The revision in reference [2] corrected a possible frameshift mutation in gag and a premature stop in env observed in the original copy of the sequence. The D194 isolate was taken from a Gambian male in 1986; the patient suffered from severe neurological symptoms up to his death in 1987.	
2 (bases 1 to 9472; revises [1]) Kuhnel, H., von Briesen, H., Dietrich, U., Adamski, M., Mix, D., Blesert, L., Kreutz, R., Immanuel, A., Hencko, K., Melchsner, C., Andresen, R., Gelderblom, H. and Rubsamen-Waigmann, H.	location/Qualifiers <5843..6138 /gene="tat" /evidence=EXPERIMENTAL join(5843..6138,8275..8371) /gene="tat" /evidence=EXPERIMENTAL /codon_start=1 /translation="METPLKEPSSLESYNEPSSCTSERDYTAQERAKOGELLAQLH RPLACNSCYCKOCYSYHCQLCFLKGLIWAYAROGRRRRPRRTKTHPPASDKSIS TRTDSOPTKKOKTPEPTIVVASGLGH" join(6069..6138,8275..8316) /gene="rev1" /evidence=EXPERIMENTAL /codon_start=1 /translation="MRDRADEEGIOELRLRLILHQINPYPHGOGTASQRNRRRRR RQMRRLVALATKLTIPDPDPSLDRAIOLOQLIQLELPDPTDIPESNSOGLAEH T" <6069..6138 /gene="rev1" /evidence=EXPERIMENTAL 1..299 /rpt_type=5' long terminal repeat /evidence=EXPERIMENTAL 1..173 /rpt_type=R repeat 5' copy /evidence=EXPERIMENTAL 59..120 /note="agtagagcctgtgtcttcctgctgtagctctcacagtcgttcgcgcgcac ctgcgcagcag in [2] revises ag in [1]" 303..319 /function="primer (lys-tRNA) binding site" /evidence=EXPERIMENTAL 547..2112 /gene="gag" /evidence=EXPERIMENTAL /codon_start=1 /translation="MGARNSYLRGKKADELEKVLRLPNGKKRYRLKHYMANELDRRI GLAESLEKSGCKRLKLEPLVPTGSENLKSLFTNYCVIMCLHAEEKYDTEAAK LAOHLVAENGTAEKMPNISRPAPSGKGNFPVOAGNYIHVPSPRLIMWVK VEEKHGAEVYPRGOLASCTECPYDINOMLNCVCDHQAAMITIELINEADAMDADQ PIPEPLDAGULDRGSDIAGTITVBOJDMWTRKOPVFPVGNITRWIDTGLCKC RMNPTNPLIDVKGPKESFQSYVDRFKSLAEQTDPAVKNMOTLILQIANDCKIK VLKQINPLIDLEMLTACOGVGSQARLAEALKEALTPAPIPFAAQQRAIRCA NCGEGHSAKQCRAPRRQGCWCKSGCHIMANCERQAGFLGMGPRGKOPNPPAAQO PGCIGTPAAPIDPAVDLLEKYMGGRRQREGRERYKEVTEDDLHLLEGCEPHRGATE DLHLNLSFGKQD" 939..940 /note="ag in [2] revises agg in [1]" 1716..4937 /gene="pol" /evidence=EXPERIMENTAL /codon_start=1 /translation="VLEIMKGGTIGETVPSTQKGLLEWQVRTHNGKLPCKTGRFFR DGPRKAPQPLRGPSSSGADNTNTPRRSSGPAVGELIYAAEKAREGEETIQGGDGG LTAPARQDAPQDRGLATPQEFILMKRPVYATLIEQDPVEVLDDGADSVIVAGIEI	



GDNYPKILVGGGFIINFKKKVNIKYULNKRVATINTGDTPIINIGRNLATLIGMS  
 LNTIPAKDILKIVTLKPKGDPRKOMPTELKTEALKEOTEKMEREGOLEAPRPNB  
 YNTPFPAKKKDKKRWMLIDPRELNRPTQDTFQLGIPRPAGLAKKRATTVLDVG  
 AIFSTPLHEDEFRQYATFLPVSNNAPERRVYKVLKPGMGSPATIQPMRQILLEP  
 KRANDVPLIEMODILIASDRTGJENHRCVVLKELNGLNGFSTPEKFEKDPFQW  
 MGYLEMPTKMKLOLPOKKEIWTYNDIOKLIVGVYMAAQIYPGIKTHLCKLRGKM  
 TLEWONTLEAELEENKIIISOEGSGYOESEELAVIKSODNOYAKTHOGE  
 RVLKGRKAKIKNTHNGVRLLAIVYOKIGELAYIVGRPKFHLPYRDIWBEOMDN  
 KYQWYVPEMDVSTPPLVRLTFNLVGPIDGETFLPYTDSGNOSKEGGAYITDKR  
 RDNRVLEQTSNQOALEAFMAALDSDSPKVNIIIVDSQYVNGIVAGIOPESENIIVN  
 ILEDMKIEAYVAVWPAHKGIGMNOEDHLVSGOIRVLELIEPAOEHESYVN  
 IKELHKGIGIPOLVARIYNTCAOQCKGEALHGVAVLEIWMQMDCTHEGKIIYVA  
 YHVAGFIEAEYIPQESRQTAFLFKLASKMPTLHTHDGPNPTSOEYKVMVAGI  
 IEOSGVYVNOGSGVVEAMNHHLKNOISREQANTETJYLAIVCHMKRKGIGIG  
 DMTPEERLIMTTDEOTFLOKRSNFKKQVYIRBGRDOLKMPGGLLKGGAIVY  
 VKVGDIVVPRKAKIIRDYGGRELDSSSHLEGAREDEVA"  
 /partial  
 4867..5514  
 /gene="y1f"  
 /evidence=EXPERIMENTAL  
 /codon\_start=1  
 /translation="MEGRNMIIVPTWVPGRMERHNSLVKHLRYTKDLEEVRYVPR  
 HKVGAWMTGSRVIFPLEGESHLEIOAWMNTPEKGMYSKSVSLTWTETKFWVPR  
 CADSLISTSTFSCPTAGEVRAIRGEELKASPMQKRNPKHROKNSLOIYLAIVVOOQ  
 RPPRGARQKMRDHRMGLRNARDQYHSLKNGGSESEPARAHPHPRYAKVLGILA"  
 5342..5580  
 /gene="vpx"  
 /evidence=EXPERIMENTAL  
 /codon\_start=1  
 /translation="MADRERYRPNCGEETIGEAFFELWDTIALNREAVNHLPREI  
 IFOWORSMAVWHDQGSTYTKRYLCIMQKAVYHFKKGCTGKRGHGGPGRMVR  
 PPPRPCLIV"  
 5680..5997  
 /gene="vpr"  
 /evidence=EXPERIMENTAL  
 /codon\_start=1  
 /translation="MTEAPTEPPEDGTPRREISWTYIETLKEIKEBALKHPRCILL  
 IALSNYIINRRGDTLEGARELIRVLFPYHIRAGCDRSKKGQTRRAPRPAFTPR  
 GMH"  
 6145..8700  
 /gene="env"  
 /evidence=EXPERIMENTAL  
 /codon\_start=1  
 /translation="MEGRNQLVAIILTSCLICYCKOYVYVYGPAMRNASIPLEFC  
 ATKNDYVTOCLPEPNDYQELNLTVEAFDAMDVYTEQALIEDVWLFESTSKPCV  
 RLTPDCAVMNOCNISGITATPSPNIIITIDENSTCIGDNNCTGKEBEVCEENMGAF  
 LEODKRKYNDAMYSRGEVCDKTNCTGTCYMRHCTSYIKESCCKHYWDKAPRYCAP  
 PGFALRNDYNSDEEPEKSGKYVAASOTBMEENOTSWPEENITBAENRYIYWHCKR  
 DNRTISLKNYINLTMHCKRPCKNTYVPTILMSGRPHSRVYKKGQAMCMQGMK  
 IEANREVQATLAKPRIGITNDGTGINTFKGDISDPYIYMNKGEELYCMTWME  
 LNWENKNTQTHGNAPCHIRQIINTMHRVGTJNVLPREDELLCNSTYISIANIDSO  
 DGNQNIIFSEAVELRYLELDGYKLEEVPIIPAPTEKERSAPYANKRGVLELFE  
 LGFATLSASGASGLTSLAOSRTLLAGIYOQOOLLDVYKROEMLRLIYWGTRKNO  
 ARVYIEVYLDQADQAINMGCAFRVCHTTPYVNDSTLTPMNNMTQDEWKRYHYLE  
 ANISLSLEADQIOQEKNNYELQKLSNDVPEGNMIDGYSWYIYQYIYVGTIGLTI  
 AIYVQLSRLKNGIRPEVSSPGYLODIHHTDROGANETEDGADGSGFELWPMW  
 PLNTYQFRIHLTLRLTLGLVNSCRGLSNKSPTRLISQSLAIRDWLRKAALAYOYG  
 CEMIOEAFARFARTARETIAGAMRGLCEBAQRIGRIGLIAYVRIRQAGAEIALT"  
 6940..6943  
 /note="gaaa in [2] revises ga in [1]"  
 6971..6973  
 /note="agg in [2] revises ag in [1]"  
 6999..7001  
 /note="caa in [2] revises ca in [1]"  
 7682..7685  
 /note="cgac in [2] revises cagc in [1]"  
 7707..7709  
 /note="cgc in [2] revises cnc in [1]"  
 8275..>8371  
 /gene="lat"  
 /evidence=EXPERIMENTAL  
 8275..8516  
 /gene="rev1"

CDS	/evidence=EXPERIMENTAL 8534..9307 /gene="nef" /evidence=EXPERIMENTAL /codon_start=1 /translation="MGASGSKRREHSQGLRERLLRARGGYVKORNASGESSESQSE GSGEKQSPSGEQOYOGGPMPTKATPAIGKNYSKOONDDVSDDDLGVGVV MPRPRLREMYTKLAIDMSHPIKEKGEGLEGFYSERRRIIDLPLEKEEGLIIPDONMT HGDPTRYPMIFGMWLMKVLPVDISOEAEEVTNCLVHPAQISRYDDEGETLVMKFDPM LAYSKAFILHPDEFGRSGDLPEKEMAKDKARCIPIYSE"
LTR	8919..9472 /rpt_type=3' long terminal repeat /evidence=EXPERIMENTAL
variation	9419..9430 /note="ggggaagccctc in [2] revises gt in [1]"
BASE COUNT	3220 a 1946 c 2382 g 1924 t
ORIGIN	
Query Match	2.2%; Score 118; DB 2; Length 9472;
Best Local Similarity	47.9%; Pred. No. 3,87e+00;
Matches	23; Conservative 10; Mismatches 15; Indels 0; Gaps 0;
Db	1081 gaaggtcgacaccctatgatcatcaatgcattagtgtgaggc 1128
Cp	1185 RAANMNCCNCNCNDATTYRTCDATNARTCTTGDTATNTNARNGC 1138
RESULT	15
LOCUS	HIVYU2 9706 bp ss-RNA VRL 19-AUG-1992
DEFINITION	Human immunodeficiency virus type 1 brain clone YU2; complete genomic sequence.
ACCESSION	M93258 M89974
SOURCE	Human Immunodeficiency virus type 1 brain clone YU2; infectious macrophage-tropic clone.
ORGANISM	Human immunodeficiency virus type 1 Viridae; ss-RNA enveloped viruses; Positive strand RNA virus; Retroviridae; Lentivirinae. 1 (bases 7005 to 7514) Li,Y., Kappeis,J.C., Conway,J.A., Price,R.W., Shaw,G.M. and Hahn,B.H.
REFERENCE	Molecular characterization of HIV-1 cloned directly from uncultured human brain: Identification or replication competent and defective viral genomes J. Virol. 65, 3973-3985 (1991)
AUTHORS	J. Virol. 65, 3973-3985 (1991) simple staff-entry 2 (bases 1 to 9706) Li,Y., Hui,H., Burgess,C.J., Price,R.W., Sharp,P.M., Hahn,B.H. and Shaw,G.M.
TITLE	Complete nucleotide sequence, genome organization, and biological properties of HIV-1 in vivo: Evidence for limited defectiveness and complementation J. Virol. (1992) In press
JOURNAL	Simple staff entry
STANDARD	
COMMENT	Kindly provided prior to publication by Shaw and Hahn, University of Alabama. YU2 clusters with HIV-1 B subtype sequences. Four full-length genomes were cloned from uncultured brain [1] and compared [2]. YU2, unlike three of the minimally defective clones (see <HIVYU10>), was found to be replication competent and macrophage tropic. 10 nonsynonymous nucleotide differences are found between YU2 and YU10, 7 of which are seen in the env cds.
FEATURES	from to/span description pept 788 2290 gag polyprotein pept 2083 5094 pol polyprotein (AA at 1) pept 5039 5617 vif protein pept 5557 5850 vpr protein pept 5831 6045 tat protein, exon 2 (first expressed exon) pept 8340 8430 tat protein, exon 3 (AA at 8341) pept 5970 6045 rev protein, exon 2 (first expressed exon) pept 8340 8614 rev protein, exon 3 (AA at 8342) pept 6062 6307 vpu protein (AA at 1) pept 6225 8756 envelope polyprotein pept 8758 9402 nef protein LTR < 1 787 5', LTR







**THIS PAGE BLANK (USPTO)**











CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.

Sequence 2424 BP; 839 A; 302 C; 462 G; 759 T;

Query Match 91.9%; Score 5012; DB 58; Length 2424;

Best Local Similarity 57.7%; Pred. No. 0.00e+00;

Matches 757; Conservative 301; Mismatches 253; Indels 0; Gaps 0;

Db 627 atgaacacatcatcttctgcaatgaatgaagttctgcatgaagttcatcaacaatc 686  
 QY 1 ATGACNCAYTATCAATTTGTGNGNATHAARGNWSNGNATWSNMTTNCNCARATH 60  
 Db 687 atgacatgattagacatgaagttcaagatcgagatattgaactgaattacagaa 746  
 QY 61 ATGCAAGAYTTGNCAGTARCTNCARGNWSNGAYATHGARAATATYNTTYNCNAR 120  
 Db 747 gtgctcttagaataaaggagataaaatattacattgctgctaatatacaataaaga 806  
 QY 121 GTNGCNTMTMGNAAYAAARGNATHAARATHYNTCCNTTGGNCAAYAAATHAARGAR 180  
 Db 807 gatatgtagtatacaagatgaatgcagcagtagcagcaggaataatgaatgca 866  
 QY 181 GAYATGGTNGTATACARGNAAAGCMTTTCGNSWNSNCAAGARARATHTGNNNGN 240  
 Db 867 catcaattgaattagattgttaagttatgaattatggttgaacagattgatca 926  
 QY 241 CAYCARNTAAATYNTGATGNTNMTNTAYAGATTTTNGNCARATHATGACAR 300  
 Db 927 tatactcagtagctgtaactggtgcacatggtlaaaactctcaacaggttatatca 986  
 QY 301 TAYACNMSGTGNCNGTNCNCGNCAAYGNAARACNWSNACNACNGNTNTNMSN 360  
 Db 987 catgtatgaatggtgataaaagaactcatttcaattgtgtgtagcagcaggtatgga 1046  
 QY 361 CAYGTNATGAAGGNGAYAAARARACNMSNTTNTTATGNGAYGNCNCGNATGGN 420  
 Db 1047 ttgctgaaagtgatattgctgcttctgagcattggaatgagcattgctcattat 1106  
 QY 421 YTNCGNARMSNGATATATTTGCTNTTTCGNCNTGTGATATATGNNGCATTTTNTMSN 480  
 Db 1107 tataaactgattacgaactatgaacaatattgtnnnnnnnnnnnnnnnnnnnnnnn 1166  
 QY 481 TAYARCCGATATAGCNATHATGACNMAATYATHGATTTGATCAVCCNCAATYATYTAAR 540  
 Db 1167 nnn 1226  
 QY 541 GAYATHAAYGATGTTTTCGATGCTNTTTCARGARATGCGCAAAATGNAARARAGNATH 600  
 Db 1227 attgctggtggtgagatgaactgaactgaactgaactgaactgaactgaactgaat 1286  
 QY 601 ATGCTGCTGGGNGAGAYARCAATYTNMGNARATGAGCNGATYTNCCNATHATYTAAR 660  
 Db 1287 tatgatttaagaatcgagatgacattatgctcaaatattcaaatatcgagataaggt 1346  
 QY 661 TAYGNGATTAAGAAVMSNGAYATHTAYGNCNARAAVATHCARATHACNCAAYAAARGN 720  
 Db 1347 actgcttttagatggtgattgagatggttattatgacactctctgctccacatat 1406  
 QY 721 AONGCNTTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
 Db 1407 gttgacatcagtttaagcattgattgattgattgattgattgattgattgattgatt 1466  
 QY 781 GGNAGATCAATAGNTNTNATAGCNTTNGCNTATGCAATHTSNTATYTNARARATYN 840  
 Db 1467 gatttatacaataatgaagaacattgattgattgattgattgattgattgattgatt 1526

QY 841 GAYTNACNAAYATHAARGNCGNTNGARACNTTTCGNGGNTNNAARMGNTTAAAT 900  
 Db 1527 gaaactaactgcaactgaactgatttgaatgattgattgattgattgattgattgatt 1586  
 QY 901 GARACNACNATHGCAACAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
 Db 1587 agtgcataactgaaacagcagcaagaataatcacataaagaagtgtgtgcagattt 1646  
 QY 961 MSNGCNACNATHGAYACNCGNMGNAARARATYCCNCAAYARGARTGNTGNCNTNTTY 1020  
 Db 1647 caaccacacacttctcagacacacagcagcattttaaattgattgctgcaagtttaagt 1706  
 QY 1021 CARCCNCAVACNTTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 Db 1707 aaagcagatcgctgattcttgaatgaattttgataactgataagaataactgagcga 1766  
 QY 1081 AARGCNGAYMGNTTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
 Db 1767 ttaacgatacaagatttaattgataaaattgaagtgatcgttaataatgaagattct 1826  
 QY 1141 YTNACNATHCARAGAYTNAHGAAYARATHTTTCGNSNATHMGARARATWSNCGN 1200  
 Db 1827 attaatgattagaacaaattgataatgctgttatttatttattggtgaggtgattt 1886  
 QY 1201 ATHAAYGNTNTGACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
 Db 1887 caaaattgataaaattgataaaattgataaaattgataaaattgataaaattgataaa 1937  
 QY 1261 CARARNTNCAARAAAGCATATTTTCATATATGATGATGATGATGATGATGATGATGAT 1311

RESULT 3  
 ID V80065 standard; DNA; 660 BP.  
 AC V80065;  
 DT 17-MAR-1999 (first entry)  
 DE Partial nucleotide sequence of the MurC gene.  
 KW MurC gene; UDP-N-acetylmuramate-L-alanine ligase; MurC polypeptide;  
 KW bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
 KW Immunogen; drug; genetic immunisation; ds.  
 OS Staphylococcus aureus.  
 FH Key Location/Qualifiers  
 FT CDS 2..660  
 FT /tag= a  
 FT /product= "partial MurC polypeptide"  
 PN EP-889123-A2.  
 PD 07-JAN-1999.  
 PF 26-JUN-1998; 305064.  
 PR 03-JUL-1997; US-052720.  
 PA (SMK ) SMITHKLINE BEECHAM CORP.  
 PA (SMK ) SMITHKLINE BEECHAM PLC.  
 PI Burnham MKR, Wallis NG;  
 DR WFT: 99-06265/06.  
 DR P-PSDB: W89199.  
 PT New isolated MurC polypeptide from Staphylococcus aureus and related  
 PT nucleic acid - useful in diagnosis, treatment and prevention of  
 PT bacterial infections  
 PS Claim 2: Pages 4-5; 39pp; English.  
 CC The invention relates to a UDP-N-acetylmuramate-L-alanine ligase  
 CC (MurC polypeptide) encoded by the S. aureus MurC gene. Host cells  
 CC containing an expression system comprising the MurC gene can be used for  
 CC the recombinant production of the polypeptide. Agonists or the MurC  
 CC polypeptide are used to treat conditions requiring increased activity or  
 CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or  
 CC competitive polypeptide are useful for inhibiting the polypeptide e.g.  
 CC bacterial (especially S. aureus) infections. They are also useful against  
 CC Helicobacter pylori infections and related cancers, ulcers and gastritis.  
 CC The antibacterial agents are useful to treat in-dwelling devices for  
 CC infection prevention or generally as wound treatments to prevent adhesion  
 CC of bacteria to matrix proteins. The MurC polypeptide is also useful for  
 CC diagnosing or prognosing a (susceptibility to) disease, for raising  
 CC antibodies, to identify modulators or specific receptors, in rational  
 CC drug design and as an immunogen for vaccines. The MurC gene sequences are  
 CC useful in antisense/ribozyme therapeutics; to detect mutant MurC gene;







KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
 RV computer readable medium; vaccine; pharmaceutical composition; ds.  
 OS Streptococcus pneumoniae.  
 PN WO9818931-A2.  
 PD 07-MAY-1998.  
 PR 30-OCT-1997; U19588.  
 PF 31-OCT-1996; US-029960.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,  
 P1 Kunsch CA, Rosen CA;  
 P2 WPI; 98-272225/24.  
 P3 Computer-readable medium with recorded Streptococcus pneumoniae  
 PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus  
 PT pneumoniae  
 PS Claim 1; Page 524-530; 1409pp; English.  
 CC The present invention describes a computer readable medium which has  
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V55254) recorded  
 CC on it, or a representative fragment or a sequence at least 95% identical  
 CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 CC to 391 (V52134 to V55254) are genomic fragments from Streptococcus  
 CC pneumoniae. The present invention also describes an isolated nucleic acid  
 CC molecule encoding a homologue of any of the fragments of the S.pneumoniae  
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 CC by a process comprising: (a) screening a genomic DNA library using as a  
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
 CC to 391, identifying members of the library which contain sequences  
 CC that hybridise to the target sequence and isolating the nucleic acid  
 CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
 CC from an organism, amplifying nucleic acid primers whose nucleotide  
 CC sequence is homologous to amplification primers derived from the  
 CC fragment of the S. pneumoniae genome to prime the amplification and  
 CC isolating the amplified sequences. The computer readable medium can be  
 CC used in a computer-based system for identifying fragments of the  
 CC S. pneumoniae genome of commercial importance, or expression modulating  
 CC fragments of the S. pneumoniae genome. Products from the present  
 CC invention can be used in diagnosis kits and assays, and pharmaceutical  
 CC compositions and vaccines for S. pneumoniae.  
 QQ Sequence 11864 BP; 3440 A; 2668 C; 2183 G; 3573 T;

Query Match	33.7%	Score 1835	DB 478	Length 11864
Best Local Similarity	42.9%	Pred. No. 1,78e-139		
Matches 550	Conservative 229	Identities 491	Indels 12	Gaps 9
D	468	atgaatattcataagtttgatggttcctgcgtccataaagcaagcaatcatgcat	527	
Cp	1279	ANGCATTGTGNARYYTTTGGDARTTCGCCNGCCCATTAANARNACNCCGRTTCRA	1220	
D	528	ggtctaggaagtcgggaaacatttccaacagtaactcgtgttttttgattttgt	587	
Cp	1219	AYTGTCNARNACRTDNTNARCTCYTTRTJDATPAAN-SWNGCNCNCQ--ATYTTT	1163	
D	588	tggtctagttcttctaccatacgtccatgatctactccacagccgacgcaataatt	647	
Cp	1162	CDATNARRTCYTGDATNTNARNCGNCNSMT--T--YTCKMDATNWMNCCRAADATYT	1106	
D	648	gcgcctagataaacagcaatctgcttggttttaagcaatgggcaagtcgctccaaagca	707	
Cp	1105	CRCANARAAANACNCKRfCNCCTYTCRCANARNSWTTCNGCRAAYTCRTNARRAANGCYT	1046	
D	708	tggtctctgtataagatcgcggttgaagaagctgtcacatttccttggttgattct	767	
Cp	1045	GNGTCKNSWRANSTRfGNGNGYfRRANACNGCNAACACTCTYTTTRfGNGGRRAYTTT	986	
D	768	gacgaagccgacatccaaagtcgcaataattctgttgatgttggtggcaagatcatga	827	
Cp	985	TNCKNGCGRTCDATNGTNCNSMDATYTCNCKNGGRTGfGNGCRRATfRCNCADA	926	
D	828	tcacgtataatgcaaatcttctccagtgaaagacggttaataaccgggcaaatcttca	887	
Cp	925	TNACTGTGTTGCDTNGTNGTTCRTTARRNCKNCKTfTfNACNCCNCCRRANfTYTCNA	866	
D	888	agtgtccagccacgaagttcaataatctctgcgtgttlaagaagaccataaagcgtg	947	

Cp	865	RNGCYTCYTTDATTTNGTGNACTCNAARITTYTCNARRRANSWDXTPNGDATTNACNGNA	806
Db	948	tcgcattcatgatataitgtagcagcaacaaggltggaatgtggaattgcccgaattgtc	10070
Cp	805	RNGCRITNARACNGTGTGTGTCNCRCRTATYTGNGSNWNNRRAARGRCRTARAAYTNC	746
Db	1008	caagggaatlgaaaggtgaaggttgaaocagttatltgaacgaaagaatcactactaca	10670
Cp	745	CRTCNACRTAATACCTCRAANGCNGTCCYTTTRCNGTDATYTGDAFTRTYTGNGCRTDA	686
Db	1068	agtcattgcctcagctctcaaacccataataaaattggtgcattcagcagcgtaacctac	11270
Cp	665	TRTCCTCNSMRTG--YTT-PAANCRTARARADYTNNGNCRFCNCGCYTDATYTTNG	629
Db	1128	gcaattcaagcctcttcacacatagacaaaagaagcccttggattgttggtagctgcgt	11870
Cp	628	KNARITGTCRTCTCCTCNCGCCANGCATDATNCCYTTTNNACRTTRIGNGCCATYTCY	569
Db	1188	taaaagcattaaacaacacctcgcagactctgttgaataactcgtatcgtgltcaaaatc	12470
Cp	568	GRANGCRITCRAAANACCTRTTDTATRTCTYTRARAFARFCNGBRGRCRAARTDART	509
Db	1248	tggtgataatagagtatctcgggtgtgtaaggaacaggaagcaccatattcgtcaagtt	13070
Cp	508	TNGTCADATATGCGCTATTCCGGYTTTRTANSWNNRRAARGNCKNCRATAYTCRANGCY	449
Db	1308	caagaacaaaatatttggcatctggcgaagcaagacacttcccatctccatcaagaagc	13670
Cp	448	CRANGCRAARTATATCNSMTTCGNNARNCCATNCNGTNCRCRTNCDDATNARRAANS	389
Db	1368	tggtatctcgtaaatgtgagaaagacaatgacacacatacctcgtctgtaagtttccat	14270
Cp	388	WNGTYTATYTTTRTCNCATTCATNACRTGNSWNNARNRNCNGTNGNSWNGYTTNCRT	329
Db	1428	gtgcctctgctactcccatcctaacaagtcacgcataaagcttaccctgaagaactca	14870
Cp	328	GNCGNCNGTWNACNGACNACMSWNGTATYATGRCODATDATTYGCNCAARRAARTCTRT	269
Db	1488	aagcttgttagcttgatataccatttltgttcgcataaagcaatttcgagcgttattcat	15470
Cp	268	ANSMNACNACRTCAARTTMAATYTGRTGNCCKNAKACDATTYCYCRT-GNSWNSWNG-C	211
Db	1548	gaaagcatttccagcgataatttccataatcaacgcgtcagatttttccatcaaaagaa	16070
Cp	210	RAANG-CRTTNCCTGTGATATACNACCATRTCTCYTTDTATRTTTCNGCNCRAANGNA	152
Db	1608	gaatgtgaattcctcgtccctcgaagacgcgcyttggtaaagtagtacttccaattc	16670
Cp	151	RDATYTTDTATNCCYTTTRTYTTCNKNARNGCNACYTCNCTRAANACPTARATYTDARTCNS	92
Db	1668	atccctgaatccctgtgcccaccatctgtgtgataaacatcaagcgaagcactcaatccctg	17270
Cp	91	WNCCYTGACCTCTGTCGNCAARRCTCGATDATYTGNCNARNNSWNSWCATNCNSWC	32
Db	1728	cctaattccgataaatgata 1749	
Cp	31	CYTTDATNCCNACRAARTGTATA 10	
RESULT 6			
ID	X13717	standard; DNA; 677 bp.	
AC	X13717		
DT	19-MAR-1999	(first entry)	
DE	Enterococcus faecalis genome contig SEQ ID NO:780.		
KW	Enterococcus faecalis; contig; detection; Enterococcal infection;		
KW	vaccine; attenuation; computer readable medium; ds.		
OS	Enterococcus faecalis.		
PN	MO985055-A2.		
PF	12-NOV-1998.		
PD	04-MAY-1998.	008985.	
PR	14-NOV-1997.	US-066009.	
RR	06-MAY-1997.	US-044031.	

RESULT	6	
ID	X13717	standard; DNA; 677 BP.
AC	X13717;	
DT	19-MAR-1999	(first entry)
DE	Enterococcus faecalis genome contig SEQ ID NO:780.	
KW	Enterococcus faecalis; contig; detection; Enterococcal infection;	
KW	vaccine; attenuation; computer readable medium; ds.	
OS	Enterococcus faecalis.	
PN	MO9850555-A2.	
PD	12-NOV-1998.	
PF	04-MAY-1998;	008985.
PR	14-NOV-1997;	US-066009.
PR	06-MAY-1997;	US-044031.



PR 16-MAY-1997; US-046655.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Dillon PJ, Kunsch CA;  
 DR WPI: 99-045171/04.  
 PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
 PT - used to develop products for the detection of Enterococcus and for  
 PT use in vaccines for prevention or attenuation of Enterococcus  
 PT infection.  
 PS Claim 1: Page 1983; 2084pp; English.  
 CC A computer readable medium has been developed which has recorded on it  
 CC 983 nucleotide sequences isolated from the Enterococcus faecalis genome.  
 CC X12938 to X12919 represent these nucleotide sequences which are primary  
 CC nucleotide sequences, also known as contigs. The computer-based system  
 CC can identify fragments of the Enterococcus faecalis genome with  
 CC commercial importance. The products can be used to detect the presence  
 CC of Enterococcus faecalis in samples. They can also be used for  
 CC diagnosing Enterococcal infection in an animal and monitoring  
 CC progression of disease, and for identifying agents which can be used to  
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
 CC another related organism, in vivo or in vitro. In particular the  
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
 CC can be used in vaccines to prevent or attenuate an Enterococcal  
 CC infection.  
 SO Sequence 677 BP; 204 A; 116 C; 142 G; 211 T;  
 Query Match 31.9%; Score 1741; DB 60; Length 677;  
 Best Local Similarity 49.0%; Pred. No. 2,62e-131;  
 Matches 329; Conservative 144; Mismatches 199; Indels 0; Gaps 0;  
 Db 6 ctgattcttcgtcattgaagcgtgtgagtcgcgcgtcattcttgcattaccag 65  
 Qy 431 SNGAYTATTGTCNTTYARCGCTGAYRTAYMGNGMCAATYTTMNSNTAYARCCNG 490  
 Db 66 atttgacattatgacgaatcgcatttgatcaccgattaccagagcattg 125  
 Qy 491 AYTAAGCATATGACNAAYTHGAYTTGAYCAICCGATATYTAARCAVTHAAYG 550  
 Db 126 acgttttttcgcgttcccaacaatgctcaccagtcacagtcacaaagaatttgc 185  
 Qy 551 AYGNNTYTGAGCMTTYTCARGARATGCGNCAYAAVGTNAARARAGNATHGCGTGG 610  
 Db 186 gtgtgtgaagatccttcgcagcagttagaaatgcaagtcagttattatcagcgta 245  
 Qy 611 GNGAYGAAGACAYTTNMGNAATHGARGCGNGAYGTCCNATHAYATYAGGNTTYA 670  
 Db 246 cgaagagagatgatatcgaagccgaataattcaacgaagcgaagccatcttg 305  
 Qy 671 ARGAYWSMGATGATATHAYGCNCARAAVTHCARATACGATYAARGNACNGCTTGG 730  
 Db 306 agtttatcacaagagatatttgtagtcatttgcctaccagcatttgccatcaca 365  
 Qy 731 AYGTNTATGTNGAYGNGARFTTYTAYGAYCTTYTWMSCNCAATYAGNGAYCA 790  
 Db 366 ataccmgaatgcgcctggtgattgctgtgcgttatttgraaactgtrtgcmaa 425  
 Qy 791 CNGNTYMAAAGCVTTCNCNTNATHGCMATWSNTATYTGARARATYAGAYTNCNA 850  
 Db 426 aagtcgcaagaaatgcttaaggttgaagcgttgaagcgttgaagcgaagaaag 485  
 Qy 851 AYTAHARAGAGCCTTNGARACNTTYGGNGNGNTMAAGMNGNTTYAAYARACNCA 910  
 Db 486 tcaatgacatgatatgttgatgatatgacgacatcgcagcgaatataaacaaga 545  
 Qy 911 THGCAAYCAAGTATGATNGAYATYAGNCAYCAACCMNGARATHTHSCNCA 970  
 Db 546 ttgatgcgcgtgcgaataataccctgacaaagaatattgctgtccttcagcacata 605  
 Qy 971 THGAYACNGCMGNAARARATYACNCAYARAGRTGTGNGCTTWTTCARCCNCAYA 1030  
 Db 606 cattacagcaaatgcttcattatgattgattgctgaagcgcctgatttgcaatg 665  
 Qy 1031 CNTTYWSNMGACNCAAGCTTYYTNAAYGARTTYGCGNGARMSYTTTGAARCCNGAYM 1090

Db 666 aagattctat 677  
 Qy 1091 GNGTWTYTYTWT 1102  
 RESULT 7  
 ID V27381 standard; DNA; 1267 BP.  
 AC V27381;  
 DT 02-OCT-1998 (first entry)  
 DE Streptococcus pneumoniae SP0070 nucleotide.  
 KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
 KW detection; pneumonia; otitis media; meningitis; ss.  
 OS Streptococcus pneumoniae.  
 FH Key location/qualifiers  
 FT CDS 2..1267  
 FT 2..1267  
 FT /"tag- a  
 FT /product= "SP0070"  
 FT /note= "no stop codon given"  
 MO9818930-A2.  
 PD 07-MAY-1998.  
 PF 30-OCT-1997; U19422.  
 PR 31-OCT-1996; US-029960.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Choi GH, Hicmocky J A, Johnson LS, Kunsch CA;  
 DR WPI: 98-272224/24.  
 DR P-PSDB: W55120.  
 PT Nucleic acid encoding antigenic peptide(s) from Streptococcus  
 PT pneumoniae - or their epitope-containing fragments, useful in  
 PT protective or therapeutic vaccines, and for diagnosis  
 PS Claim 1; Page 72; 118pp; English.  
 CC The present sequence encodes a protein from Streptococcus pneumoniae.  
 CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
 CC can be useful in vaccines for inducing protective antibodies against  
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
 CC are used to detect Streptococcus infection (by usual hybridisation or  
 CC amplification methods), also for isolating Streptococcus genes or their  
 CC allelic variants. The protein can be used similarly to detect specific  
 CC antibodies in standard immunoassays, especially for diagnosing or  
 CC monitoring infections. Antibodies which bind the protein are used to  
 CC detect corresponding antigens, to purify the protein and for passive  
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
 CC (especially 10-300) mu g/ml per dose.  
 SO Sequence 1267 BP; 365 A; 269 C; 261 G; 372 T;  
 Query Match 30.9%; Score 1687; DB 46; Length 1267;  
 Best Local Similarity 42.4%; Pred. No. 1.28e-126;  
 Matches 521; Conservative 220; Mismatches 476; Indels 12; Gaps 10;  
 Db 1 gacacatgaggagacaaagttcagggatcagatgttgaagaatgacacttaccacaag 60  
 Qy 63 GCAATGAYTTGNGNCAYGARTGNCARGGMSNGAYATHGARAATYATGNTTYACNGARGT 122  
 Db 61 cgtcttgagcagcgaagaaatccattcttccttggatgaaataaacatcagcgtga 120  
 Qy 123 NGCMTYMGNAAYARAGCNTTHAARATHYNTCCNTTYGGGNCNAAYATYTHAARGRA 182  
 Db 121 tatgaaattatcgtcgtaaatgcttcgtccagataacaaagtcgaatctcctatgc 180  
 Qy 183 YATGCTGNTATGACRGNAYAGCNTTY-G-CNMSNMSNCA-YGARARATHTGNGMNGC 239  
 Db 181 ggaacaaatgcatcgcacaaagtcacatgattcctggttcgttcattgctga 240  
 Qy 240 NCAVCARYTNAARYTNGAYGTGNTMNSNTAYAYGATYTTGCGNCARATHTHGA 299  
 Db 241 ctgttgatgagtagcagcagacacataaaacttcaagcagcagtagtctgtc 300  
 Qy 300 RTAYACNWSNGTNGCNTNACNGNGCNCAYGGNAARACNWSNACNCGNTTNTMWS 359  
 Db 301 tcatgtctgtcacaattacagataccagcttcgttgatgagatgagcagtcgctg 360  
 Qy 360 NCAVGTNATGAATGAGNGAYARARACNMSNTTYTNTATGCGNATGCGNCGNATGGG 419







```

OY 610 GGNGAIGAGCAATYTNNGNARATHGARGCNGAIGTCCNATHTATYATYAGGNTTY 669
Db 857 agcgaagaagakatalccagccc 881
OY 670 AARGAYWSNGAIGAYATHTATYAGCNC 694

RESULT 9
ID V43027 standard; DNA; 1825 BP.
AC V43027;
DE 09-NOV-1998 (first entry)
DE Streptococcus pneumoniae polypeptide coding region.
KM Polypeptide; ORF: open reading frame; Infection; bacterial;
KM Streptococcus pneumoniae; diagnosis; prophylaxis; ds.
OS Streptococcus pneumoniae.
FH Key Location/Qualifiers
FT CDS complement (256..423)
FT /tag= a
FT /note= "polypeptide"
FT /complement (731..868)
FT /tag= b
FT /note= "polypeptide"
PN M09823631-A1.
PD 04-JUN-1998.
PF 24-NOV-1997; U21976.
PR 27-NOV-1996; US-031879.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
PI Black MT, Hodson JE, Knowles DJC, Lonetto MA, Nicholas RO,
PI Reid RH, Zarfos PN;
DR WPI: 98-322654/28.
DR P-PEDB: W63755.
DR Streptococcus pneumoniae polynucleotides - useful for developing
PI products for diagnosis, prevention and treatment of infections e.g.
PI pneumonia, bacteraemia, meningitis or endocarditis
PS Claim 1: Page 151-152; 181pp; English.
PS The sequence is that of a Streptococcus polypeptide coding region.
CC The polypeptide can potentially be used for the diagnosis and
CC prevention of bacterial infections, especially SP infection.
CC It may be used for the treatment of diseases such as otitis media,
CC conjunctivitis, pneumonia, bacteraemia, meningitis, sinusitis, pleural
CC empyema, endocarditis or infection of the cerebrospinal fluid.
SQ Sequence 1825 BP; 550 A; 385 C; 366 G; 524 T;

Query Match 26.2%; Score 1428; DB 47; Length 1825;
Best Local Similarity 42.4%; Pred. No. 3,30e-104;
Matches 434; Conservative 185; Mismatches 395; Indels 10; Gaps 7;

```

```

CP 926 ATNACYTGTTCGCDATNGTNGTTCRTTANCKNCKYTTNACNCCNCRANAGTTCN 867
Db 1192 aagtgctcaagcacaagtgtaataatcctgctgtaagaagccataacgctt 1251
CP 866 ARNGCTTCTTDDATTTTNGTNCNARCTNARCTTTCNARCTTTCNARCTTTCN 807
Db 1252 gtcgcatcatalatgatactgacagcaaggttggaatgtaattgccccagttgt 1311
CP 806 ARNGCCTTNARACNGTGTGTCNCCNCRATYNGNSNANARATGTCRTARATYCN 747
Db 1312 ccaaggaatgaaggtgaaggtgaaggtgaaggtgtaaggaagaagatacctagctca 1371
CP 746 CCRTNACRTANACCTCAANCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 687
Db 1372 aagtcagctcctgacgtcgaacaaataataatgtaagtcgaagtaactcta 1431
CP 666 ATRTCNCSWRTG--YTT-RAANCCRTARATADATNGNACRTGCTGCTGCTGCTG 630
Db 1432 cgcgaatcagcatcctcacaagacaaagaccacatgtaattgttgcatagctg 1491
CP 629 CKNARCTGTCRTCTCNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 570
Db 1492 ttaagcattgaacacatcctcgaagctgtaataatcgtatgtaagtcgaagtc 1551
CP 569 TGRANCCRTCAANACCTCTTATRTCTTARATARTCNCGRTGRTCAARTCDATR 510
Db 1552 ttggtgataatagatctcgggtggtgaagcagtaagtcagtcatactgcaat 1611
CP 509 TTNGCATDATATGCTATTCNGGTTTANSMNARARATGCKNCKRTATTCRANCCY 450
Db 1612 tcaagaacaaatattgcatggtcgacacagcagctgcccataccaagaag 1671
CP 449 TCRANCCRTATATCTCNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 390
Db 1672 ctggtatctgaatgtaagcaagaagacatgacacatacctgctgtaagtttcca 1731
CP 389 SWNGYTTTTCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 330
Db 1732 tgtgtcctgctaccccatcgaagcaagtcacagcacaagtcacacacacacacag 1791
CP 329 TGNGCNCNGTNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 270
Db 1792 taac 1795
CP 269 TANS 266

RESULT 10
ID T67723 standard; DNA; 861 BP.
AC T67723;
DE 28-JUL-1997 (first entry)
DE H. pylori cytoplasmic protein ORF 11253.aa.
KM Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KM Identification; binding compound; bacterium; life cycle; activator;
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT CDS 1..861
FT /tag= a
FT /transl_except= (pos: 4..6, aa: Xaa)
FT /transl_except= (pos: 397..399, aa: Xaa)
FT /transl_except= (pos: 496..498, aa: Xaa)
FT /note= "Xaa = Unknown"
PN M09640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR ) ASTRA AB.
PI Berglindh OT, Smith D, Møllgaard BL;
DR WPI: 97-052306/05.

```



```

DR P-PSDB: 420102.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 9; Page -; 1481bp; English.
CC This sequence encodes a H. pylori cytoplasmic protein involved in
CC outer membrane or cell wall biosynthesis.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
CC Note: This DNA sequence is not reproduced in the specification and
CC has been derived from the related specification, W09719098.
SQ Sequence 861 BP; 291 A; 130 C; 169 G; 268 T;

Query March 4.1%; Score 222; DB 30; Length 861;
Matches 60; Conservative 34; Mismatches 71; Indels 0; Gaps 0;

Db 353 taagcgcttgatgaatgaatgaatgaagaatgaataatgaatgaattaaag 412
Qy 821 THWSMTAYTNGAARARTNGAYGTNMCNMAAYAHARGARGCNYTNGAACWTTGGNC 880
Db 413 gcatataaaaacgcttgatattttgcaaaaacaaatcattcatatgatgattagc 472
Qy 881 GNGTNAAMGMMNGMTTAAAYGARACNACNATHGCNAAAYCARGTATATGTGAGAYTAAG 940
Db 473 cccgcacccctactgaattgcccactttaaagaacgctagga 517
Qy 941 CNCAYCAICCNMGNGARATHWSNCCNACNATHGAYACNGCNMGA 985

RESULT 11
ID T77403 standard; DNA; 861 BP.
AC T77403;
DB 04-AUG-1997 (first entry)
DE H. pylori cytoplasmic protein ORF 11253.aa.
DT Transmembrane; cytoplasmic; cell envelope; flagella; transport;
KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
KW activator; inhibitor; bacterial life cycle; vaccine; immunise;
KW detection; antisense; inhibition; ds.
OS Helicobacter pylori.
FH Key location/Qualifiers
FT cds 1..861
FT /tag- a
FT /transl_except= (pos:4..6, aa:Xaa)
FT /transl_except= (pos:397..399, aa:Xaa)
FT /transl_except= (pos:496..498, aa:Xaa)

PN W09719098-A1.
PD 29-MAY-1997.
PF 15-NOV-1996; U18542.
PR 17-NOV-1995; US-561469.
PA (ASTR ) ASTRA AB.
PI Smith DH;
P1 WPI: 97-298052/27.
DR P-PSDB: W24585.
P-PSDB: W24585.
PT Helicobacter pylori nucleic acid sequences and related proteins -
PT used for diagnostics and therapeutics
PS Claim 1; Page 77; 235pp; English.
CC This sequence encodes an H. pylori cytoplasmic protein involved in
CC outer membrane or cell wall biosynthesis. This sequence showed
CC homology to N-acetylmuramate-Alanine ligase.
CC Helicobacter pylori has been strongly linked to chronic gastritis and
CC duodenal ulcer disease. The nucleic acid sequences of the invention

```

CC	are used to evaluate compounds, especially activators or inhibitors of
CC	bacterial life cycle, for the ability to bind an H. pylori nucleic acid
CC	sequence. The nucleic acid sequences, and corresponding proteins, are
CC	also useful for generating vaccines for immunising subjects against H.
CC	pylori or for use in detecting the presence of Helicobacter species in
CC	a sample. Antisense nucleic acid sequences of these sequences are
CC	used to inhibit expression of a gene from Helicobacter species. H.
CC	pylori whole genomic DNA was isolated and nebulised to a median size of
CC	2000 bp. Purified DNA fragments were blunt-ended and ligated to unique
CC	BstXI linker adapters in 100-1000 fold molar excess. These linkers are
CC	complementary to the BstXI-cut PMP vectors, while the overhang is not
CC	self-complementary. Therefore the linkers will not concatemise nor
CC	will the cut vector re-ligate itself easily. The linker-adaptor inserts
CC	were ligated to each of the 20 PMP vectors to construct a series of
CC	shotgun subclone libraries. The purified DNA samples were then
CC	sequenced.
CC	Note: The ORF/protein reference number for this sequence was obtained
CC	from the related specification, W09640893.
SQ	Sequence 861.BP; 291 A; 130 C; 169 G; 268 T;
Query Match	4.1%; Score 222; DB 30; Length 861;
Best Local Similarity	36.4%; Pred.No.1,44e-04;
Matches 60; Conservative 34; Mismatches 71; Indels 0; Gaps 0;	
Dd	353 taagcgttcgtagaatgaattagaagaatagaataatgatgaattgaatttaag 412
Oy	:    :    :    :    :    :    :    :    :    :    :
Dd	821 THWSTATTYTTGAAGAARTTCAGCTNCNMAATVHAARGACGNTTGACACTTGGNG 880
Oy	:    :    :    :    :    :    :    :    :    :    :
Dd	413 gcatataaaacgcgttgatatcttggcaaaaacaaccatccatccatcgatgaatag 472
Oy	:       :       :       :       :       :       :       :       :
Dd	861 GCGTNAAARMGMGTYYTAATGARCACNCAATHGCNNAAYCARGTNATHGTGATGATTAyg 940
Oy	:    :    :    :    :    :    :    :    :    :    :
Dd	473 cccacacccctactcgaattgycgccaccttaaagaacgcgatagga 517
Oy	:    :    :    :    :    :    :    :    :    :    :
Dd	941 CNCAYCAYCCNMGNRGARATHWSNCCNCAETHGAYACNGCMGMA 985
Oy	:    :    :    :    :    :    :    :    :    :    :
RESULT 12	
ID	T67859 standard; DNA; 1365 BP.
AC	T67859;
DT	14-JUL-1997 (first entry)
DE	H. pylori cytoplasmic protein ONF 0lep30520orf27.
KM	Vaccine; prevention; treatment; infection; identification;
KW	binding compound; bacterium; life cycle; activator; bacteria;
KW	inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW	cytoplasmic; outer membrane; cell wall; biosynthesis; ds.
OS	Helicobacter pylori.
FH	Key Location/Qualifiers
FT	cds 1..1365
FT	/tag= "a
FT	/note= "no stop codon given"
PN	W09640893-A1.
PD	19-DEC-1996
PF	06-JUN-1996; U09122:
PR	07-JUN-1995; US-487032.
PR	01-APR-1996; US-630405.
PA	(ASTR ) ASTRA AB.
P1	Berglundh OT, Smith D, Mellgaard BL;
DR	WP1: 97-052306/05.
DR	P-PDSB: W20606.
PT	Helicobacter pylori nucleic acid sequences and related
PT	polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT	infection, and to detect Helicobacter
PS	Claim 9; Page 752; 1481pp; English.
CC	The present sequence encodes a Helicobacter pylori cytoplasmic
CC	protein involved outer membrane or cell wall biosynthesis.
CC	The protein may be used in a vaccine to prevent or treat H. pylori
CC	infection or to identify H. pylori polypeptide binding compounds,
CC	useful as potential H. pylori life cycle activators or inhibitors.
CC	The genomic sequence of H. pylori (ATCC 55679) was determined from
CC	overlapping conigs generated by mechanically shearing the
CC	bacterial DNA. The sequences were analysed for ORF of at least 180
CC	nucleotides and the predicted coding regions derived by computer











**THIS PAGE BLANK (USPTO)**





(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_npn n.a. - n.a. Smith-Waterman search, using a protein query  
which has been backtranslated into n.a. using IUPAC symbols

Run on: Sat Nov 27 09:30:52 1999; Maspar time 96.53 Seconds  
1174.641 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-2  
Description: (1-437) from US09103287.pep  
Perfect Score: 5451  
N.A. Sequence: 1 ATGACGATATCAATTCCTG.....TNGNATGAARAAGCCTTY 1311  
Comp: TACTGTGATGCTTATCAACA.....ANCNTACTTTTCGNAAR

Scoring table: TABLE bkttranslate2  
Gap 30

Nmatch STD : Dbase 0; Query 0

Searched: 165359 seqs, 43243793 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: n-issued  
1:5A\_COMB 2:5B\_COMB 3:5C\_COMB 4:PCT9\_COMB 5:backfiles1

Statistics: Mean 51.193; Variance 211.724; scale 0.242

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
C 1	319	5.9	7218	2	US-08-232-	Sequence 14, Applicatio	3.61e-15
C 2	183	3.4	1350	2	US-08-665-	Sequence 1, Applicatio	1.41e-03
C 3	183	3.4	1350	2	US-08-665-	Sequence 1, Applicatio	1.41e-03
C 4	183	3.4	1350	3	US-08-843-	Sequence 3, Applicatio	1.41e-03
C 5	183	3.4	1353	3	US-08-843-	Sequence 1, Applicatio	1.41e-03
C 6	158	2.9	4732	5	5521093-4	Patent No. 5521093.	1.18e-01
C 7	147	2.7	2380	3	US-08-572-	Sequence 3, Applicatio	7.64e-01
C 8	143	2.6	2277	3	US-08-676-	Sequence 2, Applicatio	1.48e+00
C 9	143	2.6	2277	4	PCT-US91-0	Sequence 1, Applicatio	1.75e+00
C 10	142	2.6	2588	4	PCT-US91-0	Sequence 1, Applicatio	1.75e+00
C 11	142	2.6	2608	1	US-07-725-	Sequence 1, Applicatio	1.75e+00
C 12	139	2.5	401	1	US-08-137-	Sequence 3, Applicatio	2.86e+00
C 13	137	2.5	1659	1	US-08-231-	Sequence 1, Applicatio	3.97e+00
C 14	137	2.5	1659	4	PCT-US93-0	Sequence 10, Applicati	5.48e+00
C 15	135	2.5	4766	4	PCT-US91-0	Sequence 18, Applicati	7.55e+00
C 16	133	2.4	540	2	US-08-117-	Sequence 1, Applicatio	1.75e+00
C 17	129	2.4	774	4	PCT-US91-0	Sequence 59, Applicati	7.55e+00
C 18	133	2.4	1250	2	US-08-117-	Sequence 1, Applicatio	1.75e+00
C 19	132	2.4	1659	1	US-08-231-	Sequence 2, Applicatio	8.86e+00

C 20	132	2.4	1659	1	US-08-231-	Sequence 1, Applicatio	8.86e+00
C 21	132	2.4	1754	2	US-08-507-	Sequence 2, Applicatio	8.86e+00
C 22	129	2.4	8585	1	US-08-030-	Sequence 3, Applicatio	1.42e+01
C 23	128	2.3	774	4	PCT-US91-0	Sequence 1, Applicatio	1.67e+01
C 24	128	2.3	1110	1	US-08-229-	Sequence 57, Applicatio	1.67e+01
C 25	128	2.3	1110	1	US-08-630-	Sequence 11, Applicati	1.67e+01
C 26	127	2.3	1428	4	PCT-US94-0	Sequence 15, Applicatio	1.95e+01
C 27	127	2.3	1515	4	PCT-US94-0	Sequence 3, Applicatio	2.28e+01
C 28	126	2.3	1642	2	US-08-426-	Sequence 3, Applicatio	2.28e+01
C 29	126	2.3	1642	2	US-08-401-	Sequence 3, Applicatio	2.28e+01
C 30	126	2.3	1642	1	US-08-354-	Sequence 3, Applicatio	2.28e+01
C 31	126	2.3	1642	1	US-07-999-	Sequence 3, Applicatio	2.28e+01
C 32	126	2.3	1642	1	US-08-426-	Sequence 1, Applicatio	2.28e+01
C 33	127	2.3	1747	1	US-08-486-	Sequence 1, Applicatio	2.28e+01
C 34	125	2.3	2277	3	US-08-676-	Sequence 2, Applicatio	2.66e+01
C 35	125	2.3	2277	2	US-08-576-	Sequence 2, Applicatio	2.66e+01
C 36	126	2.3	2302	1	US-08-354-	Sequence 1, Applicatio	2.28e+01
C 37	126	2.3	2302	2	US-08-426-	Sequence 1, Applicatio	2.28e+01
C 38	126	2.3	2302	1	US-07-999-	Sequence 1, Applicatio	2.28e+01
C 39	126	2.3	2302	1	US-08-426-	Sequence 1, Applicatio	2.28e+01
C 40	126	2.3	2302	2	US-08-401-	Sequence 1, Applicatio	1.95e+01
C 41	127	2.3	2522	1	US-08-559-	Sequence 1, Applicatio	1.95e+01
C 42	125	2.3	3000	1	US-07-841-	Sequence 3, Applicatio	2.66e+01
C 43	125	2.3	3000	3	US-08-290-	Sequence 3, Applicatio	2.66e+01
C 44	125	2.3	5981	3	US-08-290-	Sequence 83, Applicati	2.66e+01
C 45	128	2.3	7785	3	US-08-276-	Sequence 1, Applicatio	1.67e+01

## ALIGNMENTS

RESULT 1  
ID US-08-232-463-14 STANDARD: DNA; UNC; 7218 BP.  
AC xxxxxx  
DE Sequence 14, Application US/08232463  
CC Sequence 14, Application US/08232463  
CC Patent No. 5670367  
CC GENERAL INFORMATION:  
CC APPLICANT: DORNER, F.  
CC APPLICANT: SCHEIFLINGER, F.  
CC APPLICANT: FALKNER, F. G.  
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
CC NUMBER OF SEQUENCES: 52  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Foley & Lardner  
CC STREET: 1800 Diagonal Road, Suite 500  
CC CITY: Alexandria  
CC STATE: VA  
CC COUNTRY: USA  
CC ZIP: 22313-0299  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: IBM PC compatible  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/232,463  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/935,313  
CC FILING DATE:  
CC APPLICATION NUMBER: EP 91 114 300.6  
CC FILING DATE: 25-AUG-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: BENT, Stephen A.  
CC REGISTRATION NUMBER: 29,768  
CC REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (703)836-9300  
CC TELEFAX: (703)836-4109  
CC TELEEX: 899149  
CC INFORMATION FOR SEQ ID NO: 14:



```

CC ATTORNEY/AGENT INFORMATION:
CC NAME: Webster, Thomas D
CC REGISTRATION NUMBER: 39,872
CC REFERENCE/DOCKET NUMBER: X-9900
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 317-276-3334
CC TELEFAX: 317-276-3861
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1350 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: mRNA
CC HYPOTHEICAL: NO
CC ANTI-SENSE: NO
SQ SEQUENCE 1350 BP: 384 A; 272 C; 328 G; 0 T; 366 OTHER.

Query Match 3.4%; Score 183; DB 2; Length 1350;
Best Local Similarity 29.5%; Pred. No. 1,41e-03;
Matches 33; Conservative 32; Mismatches 47; Indels 0; Gaps 0

Db 831 CCACAUGGGAAGGAGCCGCGAGCUAUGGCGUAGCCAGCUCUGGAGUGGACAA 890
Oy 786 YCAACAGCTNTNNAAGCTNTGCGTNGCTGATGCAATGMSNTAYTTNGARARITNGAYGT 845
Db 891 UCAACCAUGCAAGGAACUCUCUUCAGCCUCUCUGUGUGUGUCCAACACCGCUC 942
Oy 846 NACNAAYATHAARGARCGCTNGARACCTTYGGNGGCTNNAARMGNGMTTY 897

RESULT 3
ID US-08-665-435A-1 STANDARD; DNA; UNC; 1350 BP.
AC xxxxxx
DE
DT
Sequence 1, Application US/08665435A
Sequence 1, Application US/08665435A
Patent No. 5681694
GENERAL INFORMATION:
CC APPLICANT: Skatrud, Paul
CC APPLICANT: Peery, Robert
CC APPLICANT: Hoskins, Joan
CC APPLICANT: Wu, Chyun-Yeh Earnest
CC TITLE OF INVENTION: Biosynthetic Gene Muid of Streptococcus
CC NUMBER OF SEQUENCES: 3
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Eli Lilly and Company
CC STREET: Lilly Corporate Center
CC CITY: Indianapolis
CC STATE: Indiana
CC COUNTRY: US
CC ZIP: 46285
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/665,435A
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Webster, Thomas D
CC REGISTRATION NUMBER: 39,872
CC REFERENCE/DOCKET NUMBER: X-9900
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 317-276-3334
CC TELEFAX: 317-276-3861
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1350 base pairs
CC TYPE: nucleic acid
CC

```







```

CC APPLICATION NUMBER: US/08/572,951
CC FILING DATE: 15-DEC-1995
CC CLASSIFICATION: 800
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/346,602
CC FILING DATE: 29-NOV-1994
CC CLASSIFICATION: 800
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/263,921
CC FILING DATE: 21-JUN-1994
CC CLASSIFICATION: 800
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Paul N. Kokulis
CC REGISTRATION NUMBER: 16,773
CC REFERENCE/DOCKET NUMBER: 222957/1.02.15C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202) 861-3000
CC TELEFAX: (202) 822-0944
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 2380 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: both
CC TOPOLOGY: linear
CC MOLECULE TYPE: CDNA
CC SEQUENCE 2380 BP: 333 A; 266 C; 518 G; 272 T; 991 OTHER.
SO

Query Match 2.7%; Score 147; DB 3; Length 2380;
Best Local Similarity 14.5%; Pred. No. 7,64e-01;
Matches 16; Conservative 43; Mismatches 50; Indels 1; Gaps 1

Db 1482 YTTNCAWSSNAGAYATVACNAAYTAVACNTTYGARACNTNGAYACNGNARMGCA 1541
Cc 571 CYTGAAGCGCTCRAACACRCFTTDAATRCYTTTAAATATCTCGGR-TGTCGAATCD 513

Db 1542 RTGYAARCCNCNTNCARMGNCARYTNGNGYTTGCATGTMNGAYATG 1591
Cc 512 AATTNGCATDATNGCTTARTCTCGGTTTTRANSWMARRAARTGNCCK 463

RESULT 8
ID US-08-676-974-2 STANDARD; DNA; UNC; 2277 BP.
AC xxxxxx

Dt Sequence 2, Application US/08676974
Cc Sequence 2, Application US/08676974
Cc Patent No. 5770422
Cc GENERAL INFORMATION:
Cc APPLICANT: COLLINS, KATHLEEN
Cc TITLE OF INVENTION: Human Telomerase
Cc NUMBER OF SEQUENCES: 10
Cc CORRESPONDENCE ADDRESS:
Cc ADDRESSEE: Science & Technology Law Group
Cc STREET: 268 Bush Street, Suite 3200
Cc CITY: San Francisco
Cc STATE: CA
Cc COUNTRY: USA
Cc ZIP: 94104
Cc COMPUTER READABLE FORM:
Cc MEDIUM TYPE: Floppy disk
Cc COMPUTER: IBM PC compatible
Cc OPERATING SYSTEM: PC-DOS/MS-DOS
Cc SOFTWARE: PatentIn Release #1.0, Version #1.30
Cc CURRENT APPLICATION DATA:
Cc APPLICATION NUMBER: US/08/676,974
Cc FILING DATE:
Cc CLASSIFICATION: 530
Cc ATTORNEY/AGENT INFORMATION:
Cc NAME: Osman Ph.D., Richard A
Cc REGISTRATION NUMBER: 36,627
Cc REFERENCE/DOCKET NUMBER: UCB96-055
Cc TELECOMMUNICATION INFORMATION:
Cc TELEPHONE: (415)343-4341

```



Cc	TELEFAX:	(415)343-4342
Cc	INFORMATION FOR SEQ ID NO:	2:
Cc	SEQUENCE CHARACTERISTICS:	
Cc	LENGTH:	2277 base pairs
Cc	TYPE:	nucleic acid
Cc	STRANDEDNESS:	double
Cc	TOPOLOGY:	linear
Cc	MOLECULE TYPE:	CDNA
SQ	SEQUENCE	2277 BP; 511 A; 212 C; 395 G; 216 T; 943 OTHER.
Dd	Query Match	2.6%; Score 143; DB 3; Length 2277;
Dd	Best Local Similarity	14.4%; Pred. No. 1,486+00;
Cp	Matches	27; Conservative 63; Mismatches 97; Indels 0; Gaps 0;
Dd	663	TGGARGANGARARAAYGAYGAYGAYGAYGAYGAYGARGANGAGGNGTNTTYG 742    ::: :
Cp	569	TGRANCGRTCRANAACRTRCTDATTTCYTTTRAARTFGTNGGRTGTCRAATCATR 510
Dd	743	AYGAGGAGCAGGARBARABARAAIATHGARWSNAAAGTMACNAACCGTNCARATHC 802 ::: :
Cp	509	TTNGCATADATNGCCTATTCCGTGTTTFANSWNAARRAARTGCNCKNRATYATCAGCY 450
Dd	803	ARAAAMGCGMGTNARNRGNCNCGCCNCAARWSMSNGAYCAWSNGAGTAAATG 862 :
Cp	449	TGRANCGRRARTARTCTSWITTCNGNARNRCCCATNCCNGTNCNCDCATNARAA 390
Dd	863	SNGAYYT 869   :     :
Cp	389	SMNGTYT 383
RESULT	9	
ID	US-08-676-967-2 STANDARD; DNA; UNC; 2277 BP.	
AC	xxxxxx	
DE		
DT		
Sequence 2,	Application US/08676967	
Sequence 2,	Application US/08676967	
Patent No.	5/47317	
GENERAL INFORMATION:		
APPLICANT:	COLLINS, KATHLEEN	
TITLE OF INVENTION:	Human Telomerase	
NUMBER OF SEQUENCES:	10	
CORRESPONDENCE ADDRESS:		
ADDRESSEE:	Science & Technology Law Group	
STREET:	268 Bush Street, Suite 3200	
CITY:	San Francisco	
STATE:	CA	
COUNTRY:	USA	
ZIP:	94104	
COMPUTER READABLE FORM:		
MEDIUM TYPE:	Floppy disk	
COMPUTER:	IBM PC compatible	
OPERATING SYSTEM:	PC-DOS/MS-DOS	
SOFTWARE:	Patentin Release #1.0, Version #1.30	
CURRENT APPLICATION DATA:		
APPLICATION NUMBER:	US/08/676,967	
FILING DATE:		
CLASSIFICATION:	530	
ATTORNEY/AGENT INFORMATION:		
NAME:	Osmar Ph.D., Richard A	
REGISTRATION NUMBER:	36,627	
REFERENCE/DOCKET NUMBER:	UCB96-055	
TELECOMMUNICATION INFORMATION:		
TELEPHONE:	(415)343-4341	
TELEFAX:	(415)343-4342	
INFORMATION FOR SEQ ID NO:	2:	
SEQUENCE CHARACTERISTICS:		
LENGTH:	2277 base pairs	
TYPE:	nucleic acid	
STRANDEDNESS:	double	
TOPOLOGY:	linear	
MOLECULE TYPE:	CDNA	
SEQUENCE	2277 BP; 511 A; 212 C; 395 G; 216 T; 943 OTHER.	

[illegible]







CP 260 ACRTGNARYTTNARYTGTGTGNGCCKNACDATTCTCTCGTGN 218

RESULT 13

ID US-08-231-729B-2 STANDARD; DNA; UNC; 1659 BP.  
AC xxxxxx  
DE Sequence 2, Application US/08231729B  
CC Sequence 2, Application US/08231729B  
CC Patent No. 5618722

GENERAL INFORMATION:

CC APPLICANT: ZENNO, Shuhei  
CC APPLICANT: SHIRAIISHI, Shinji  
CC APPLICANT: INOUE, Satoshi  
CC APPLICANT: SAIGO, Kaoru  
CC TITLE OF INVENTION: FIREFLY LUCIFERASE GENE  
CC NUMBER OF SEQUENCES: 8  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: LEYDIG, VOIT & MAYER  
CC STREET: 700 Thirteenth Street, N.W., Suite 300  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20005

COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/231.729B  
CC FILING DATE: 20-APR-1994  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: JP 119050/1993  
CC FILING DATE: 21-APR-1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Rose, Herbert C.  
CC REGISTRATION NUMBER: 29846  
CC REFERENCE/DOCKET NUMBER: 60130/No. 5618722aka  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 202-737-6770  
CC TELEFAX: 202-737-6776  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1659 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: CDNA  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 1..1659  
CC OTHER INFORMATION: for codons 28, 32, 112, 130, 142, 190, 212,  
CC OTHER INFORMATION: 217, 222, 266, 329, 336, 436, 512, and 532, if the 3'  
CC OTHER INFORMATION: nucleotide is T or C, then the 5' nucleotide is C; and if  
CC OTHER INFORMATION: nucleotide is A, then the 3' nucleotide is A or G  
SQ SEQUENCE 1659 BP; 321 A; 193 C; 268 G; 265 T; 612 OTHER.

Query Match 2.5%; Score 137; DB 1; Length 1659;  
Best Local Similarity 21.3%; Pred. No. 3.97e+00;

Matches 17; Conservative 30; Mismatches 33; Indels 0; Gaps 0;

DB 1576 ACNGNARATGAYACNMGNAARATHTATNATHAAGCNCARAGGNAAR 1635

QY 674 AYMSNGAYGAYATHTATGACNARATHTATNATHAAGCNCARAGGNAAR 733

DB 1636 TCNARTCNARAGGNAAR 1655

QY 734 TMTAYGTGAGGNGARTTY 753

RESULT 14  
ID US-08-231-729B-1 STANDARD; DNA; UNC; 1659 BP.  
AC xxxxxx

DE Sequence 1, Application US/08231729B  
CC Sequence 1, Application US/08231729B  
CC Patent No. 5618722

GENERAL INFORMATION:

CC APPLICANT: ZENNO, Shuhei  
CC APPLICANT: SHIRAIISHI, Shinji  
CC APPLICANT: INOUE, Satoshi  
CC APPLICANT: SAIGO, Kaoru  
CC TITLE OF INVENTION: FIREFLY LUCIFERASE GENE  
CC NUMBER OF SEQUENCES: 8  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: LEYDIG, VOIT & MAYER  
CC STREET: 700 Thirteenth Street, N.W., Suite 300  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20005

COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/231.729B  
CC FILING DATE: 20-APR-1994  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: JP 119050/1993  
CC FILING DATE: 21-APR-1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Rose, Herbert C.  
CC REGISTRATION NUMBER: 29846  
CC REFERENCE/DOCKET NUMBER: 60130/No. 5618722aka  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 202-737-6770  
CC TELEFAX: 202-737-6776  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1659 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: CDNA  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 1..1659  
CC OTHER INFORMATION: "Xaa" at codon 409 is either Glu or Asp.  
CC OTHER INFORMATION: "Xaa" at codons 28, 32, 112, 130, 142, 190, 212, 217, 2  
CC OTHER INFORMATION: 329, 336, 386, 436, 512, and 532 is either Arg, Ser or  
SQ SEQUENCE 1659 BP; 321 A; 193 C; 268 G; 265 T; 612 OTHER.

Query Match 2.5%; Score 137; DB 1; Length 1659;  
Best Local Similarity 21.3%; Pred. No. 3.97e+00;

Matches 17; Conservative 30; Mismatches 33; Indels 0; Gaps 0;

DB 1576 ACNGNARATGAYACNMGNAARATHTATNATHAAGCNCARAGGNAAR 1635

QY 674 AYMSNGAYGAYATHTATGACNARATHTATNATHAAGCNCARAGGNAAR 733

DB 1636 TCNARTCNARAGGNAAR 1655

QY 734 TMTAYGTGAGGNGARTTY 753

RESULT 15

ID PCT-US93-07261-10 STANDARD; DNA; UNC; 4766 BP.  
AC xxxxxx

DE Sequence 10, Application PC/TUS9307261



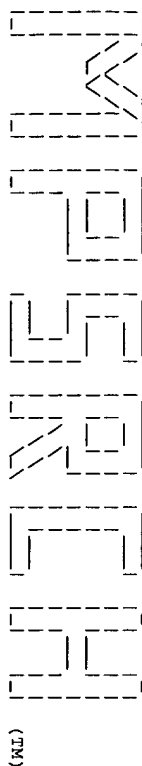
```

CC Sequence 0, Application PC/TUS9307261
CC GENERAL INFORMATION:
CC TITLE OF INVENTION: PfEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREC
CC NUMBER OF SEQUENCES: 23
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: John H. C. Blasdale
CC STREET: One Giralda Farms
CC CITY: Madison
CC STATE: New Jersey
CC COUNTRY: USA
CC ZIP: 07940-1000
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: Apple Macintosh
CC OPERATING SYSTEM: Macintosh 6.0.5
CC SOFTWARE: Microsoft Word 5.1a
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/07261
CC FILING DATE: 19930805
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/927,531
CC FILING DATE: 07-AUG-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Blasdale, John H. C.
CC REGISTRATION NUMBER: 31,895
CC REFERENCE/DOCKET NUMBER: DX0288K
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 201-822-7398
CC TELEFAX: 201-822-7039
CC INFORMATION FOR SEQ ID NO: 10:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 4766 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: cdna to mRNA
CC ORGANISM: Plasmodium falciparum
CC STRAIN: Malayan Camp
CC IMMEDIATE SOURCE:
CC CLONE: p2b1:p12-1
CC FEATURE:
CC NAME/KEY: CDS
CC LOCATION: 3..4766
SQ SEQUENCE 4766 BP: 2403 A; 508 C; 863 G; 992 T; 0 OTHER.
Query Match 2.5%; Score 135; DB 4; Length 4766;
Best Local Similarity 37.0%; Pred. No. 5,48e+00;
Matches 51; Conservative 26; Mismatches 61; Indels 0; Gaps 0;
Db 583 AACGTGAATGAAAGATGAAGTGGATGATGAATGATTAAAAAATATGATGATGAACACGAG 642
Oy 620 ARCAYYTTMGNNAARHAGRCNGCAVGTNCCNATHHTATATATAGGNTTYAARGAWMSG 679
Db 643 AAAAGAATGTCRACGAAGATGACAGAAATTAAGATTAAGGTGATGCGTTATGACAGAAATTG 702
Oy 680 AYGAATHTATVGCNCARRAAYATHCARATHACGAYAAAGNACMCNTTYGAYGTNTAYG 739
Db 703 TTGAACAACAATTTATG 720
Oy 740 TNGAATGNGARTTYTAG 757

```

Search completed: Sat Nov 27 09:32:33 1999  
Job time : 101 secs.





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Nov 24 00:36:34 1999; Maspar time 19.53 Seconds

Tabular output not generated. 896.452 Million cell updates/sec

Title: >US-09-103-287-2  
Description: (1-437) from US09103287.pep  
Perfect Score: 3121  
Sequence: 1 MTHYFVGIGSGMSLAQI.....GDIOKLNAYLDKGMKNAF 437

Scoring table: PAM 150  
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r60  
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 50.576; Variance 122.235; scale 0.414

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1975	63.3	432	2	C69662	UDP-N-acetylmuramate-	0.00e+00
2	620	19.9	434	2	B70418	UDP-N-acetylmuramate-	2.69e-77
3	602	19.3	481	2	C71338	Probable UDP-N-acetyl	2.58e-74
4	584	18.7	495	2	C71679	UDP-N-acetylmuramate-	2.42e-71
5	581	18.6	468	2	H70201	UDP-N-acetylmuramate-	7.56e-71
6	567	18.2	803	2	A71475	Probable muramate-Ala	1.53e-68
7	529	16.9	475	2	E64185	UDP-N-acetylmuramate-	2.60e-62
8	495	15.9	491	1	CEECAM	UDP-N-acetylmuramate-	8.97e-57
9	470	15.1	505	2	B67022	UDP-N-acetylmuramoyl	9.98e-53
10	440	14.1	453	2	B46002	hypothetical protein	6.64e-48
11	424	13.6	457	2	S56459	hypothetical 48.5K pr	2.39e-45
12	422	13.5	449	2	G64597	UDP-N-acetylmuramate-	4.98e-45
13	409	13.1	449	2	B71917	UDP-N-acetylmuramate-	5.81e-43
14	349	11.2	494	2	D70579	Probable murC protein	1.52e-33
15	180	5.8	494	2	B47691	UDP-N-acetylmuramoyl	7.76e-09
16	177	5.7	457	2	E71699	UDP-N-acetylmuramoyl	1.93e-08
17	160	5.1	445	2	A71659	UDP-N-acetylmuramoyl	3.05e-06
18	158	5.1	505	2	S75968	UDP-N-acetylmuramoyl	5.45e-06
19	154	4.9	682	2	A69170	UDP-N-acetylmuramyl t	1.73e-05
20	147	4.7	450	2	B69198	UDP-N-acetylmuramyl t	1.27e-04
21	142	4.5	508	2	A70125	UDP-N-acetylmuramyl t	5.11e-04
22	128	4.1	445	2	G70371	UDP-MURNAc-pentapepti	2.23e-02
23	128	4.1	449	2	JC6560	UDP-N-acetylmuramoyl	2.23e-02

24	128	4.1	457	2	F69662	UDP-N-acetylmuramoyl	2.23e-02
25	124	4.0	856	2	A64699	hypothetical protein	6.32e-02
26	121	3.5	464	2	H70137	UDP-N-acetylmuramoyl	1.36e-01
27	118	3.8	451	2	D47691	UDP-N-acetylmuramoyl	2.90e-01
28	119	3.8	482	2	E70450	UDP-MurNAc-tripeptide	2.26e-01
29	120	3.8	598	2	D71127	hypothetical protein	1.75e-01
30	116	3.7	437	2	B64185	UDP-N-acetylmuramoyl	4.77e-01
31	111	3.6	121	2	S18067	thyroid hormone recep	1.62e+00
32	113	3.6	320	2	S33172	ornithine cyclodamin	9.96e-01
33	111	3.6	432	2	H64383	Na+ transporter - Met	1.62e+00
34	111	3.6	493	2	D64612	UDP-MurNAc-pentapepti	1.62e+00
35	108	3.5	143	2	E69065	molybdenum cofactor b	3.30e+00
36	109	3.5	452	2	F64730	UDP-N-acetylmuramoyl	2.61e+00
37	109	3.5	585	2	C70330	conserved hypothetical	2.61e+00
38	110	3.5	1558	2	B71603	RESA-H3 antigen PEB09	2.05e+00
39	110	3.5	1670	2	S71551	DNA-directed DNA poly	2.05e+00
40	105	3.4	227	2	H64336	formate dehydrogenase	6.65e+00
41	106	3.4	374	2	S75646	hypothetical protein	5.28e+00
42	106	3.4	416	2	E69599	competence-damage ind	5.28e+00
43	106	3.4	447	2	D71812	UDP-N-acetylmuramyl-t	5.28e+00
44	106	3.4	447	2	F64706	UDP-MurNAc-tripeptide	5.28e+00
45	105	3.4	758	2	S37855	hypothetical protein	6.65e+00

#### ALIGNMENTS

RESULT	1
ENTRY	C69662
TITLE	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -
ORGANISM	Bacillus subtilis
DATE	#formal_name Bacillus subtilis
DATE	05-Dec-1997
DATE	#sequence_revision 05-Dec-1997 #text_change 17-Mar-1999
ACCESSIONS	C69662; S71002
REFERENCE	A69580
#authors	Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolyard, A.; Borchert, S.; Bortiss, R.; Boursier, L.; Brans, A.; Braun, M.; Briganello, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Caporaso, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, S.A.; Denizot, F.; Devine, K.M.; Diesterheft, A.; Ehrlich, S.D.; Emerson, P.T.; Etlinger, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galloni, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, I.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maneel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Mostl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelie, D.; Porwili, S.; Prescott, A.M.; Presenc, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Seror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassier, A.; Viari, A.; Wambuit, R.; Wedler, E.; Wedler, H.; Wetzinger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

#journal  
#title  
The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.



Accession	Gene	Protein	Accession	Gene	Protein
#cross-references MUID:98044033			#cross-references MUID:98044033		
#accession C69662			#accession C69662		
#status	preliminary: nucleic acid sequence not shown;		#status	preliminary: nucleic acid sequence not shown;	
	translation not shown			translation not shown	
##molecule_type DNA			##molecule_type DNA		
##residues 1-432 ##label KUN			##residues 1-432 ##label KUN		
##cross-references GB:299119; GN:AL009126; NID:92635411; PID:cell185852;			##cross-references GB:299119; GN:AL009126; NID:92635411; PID:cell185852;		
##experimental_source strain 168			##experimental_source strain 168		
##accession S71000			##accession S71000		
Varon, D.; Brody, M.S.; Price, C.W.			Varon, D.; Brody, M.S.; Price, C.W.		
Mol. Microbiol. (1996) 20:339-350			Mol. Microbiol. (1996) 20:339-350		
Bacillus subtilis operon under the dual control of the			Bacillus subtilis operon under the dual control of the		
general stress transcription factor sigma(B) and the			general stress transcription factor sigma(B) and the		
sporulation transcription factor sigma(H).			sporulation transcription factor sigma(H).		
#cross-references MUID:96310371			#cross-references MUID:96310371		
##accession S71002			##accession S71002		
##molecule_type DNA			##molecule_type DNA		
##residues 85-432 ##label VAR			##residues 85-432 ##label VAR		
##cross-references EMBL:L31845; NID:9556013; PID:9556014			##cross-references EMBL:L31845; NID:9556013; PID:9556014		
##experimental_source strain 168, substrain Marburg			##experimental_source strain 168, substrain Marburg		
GENETICS			GENETICS		
gene muc			gene muc		
FUNCTION			FUNCTION		
pathway peptidoglycan biosynthesis			pathway peptidoglycan biosynthesis		
keywords ATP binding; cell division; cell wall; ligase; peptidoglycan biosynthesis			keywords ATP binding; cell division; cell wall; ligase; peptidoglycan biosynthesis		
SUMMARY			SUMMARY		
Query Match 63.3%; Score 1975; DB 2; Length 432;			Query Match 63.3%; Score 1975; DB 2; Length 432;		
Best Local Similarity 63.9%; Pred.No.0.00e+00;			Best Local Similarity 63.9%; Pred.No.0.00e+00;		
Matches 273; Conservative 60; Mismatches 94; Indels 0; Gaps 0;			Matches 273; Conservative 60; Mismatches 94; Indels 0; Gaps 0;		
1 MTGVVHFVGIKGTGMSPLAQIHLHDNGYTVQSGDIEKFTQTQALEKRNITITLPEFSAENIKP 60			1 MTGVVHFVGIKGTGMSPLAQIHLHDNGYTVQSGDIEKFTQTQALEKRNITITLPEFSAENIKP 60		
1 MTHHYFVGIKSGMSGLAQIMHDLGHEVQSDIENYVFTFVALNKKIKITLPGANNIKE 60			1 MTHHYFVGIKSGMSGLAQIMHDLGHEVQSDIENYVFTFVALNKKIKITLPGANNIKE 60		
61 GNTVIAGNAFPDTHPELEKAMSEGIPIYIRYHKKFLGDMYKFTSVAAVTGAHGTSTGLLA 120			61 GNTVIAGNAFPDTHPELEKAMSEGIPIYIRYHKKFLGDMYKFTSVAAVTGAHGTSTGLLA 120		
61 DMVVIQGNAPFASHSHEIYRAHQKLIDVYSYNDPFGQIIDQYTSVAVTGAGKGTSTGLLS 120			61 DMVVIQGNAPFASHSHEIYRAHQKLIDVYSYNDPFGQIIDQYTSVAVTGAGKGTSTGLLS 120		
121 HVIQNAKRTSFLIGDGTQGSNENSEYFVFEKCEKRRHFLSQRPYALMTNIDDPHYES 180			121 HVIQNAKRTSFLIGDGTQGSNENSEYFVFEKCEKRRHFLSQRPYALMTNIDDPHYES 180		
121 HYMMGDKRTSFLIGDGTQGSNENSEYFVFEKCEKRRHFLSQRPYALMTNIDDPHYES 180			121 HYMMGDKRTSFLIGDGTQGSNENSEYFVFEKCEKRRHFLSQRPYALMTNIDDPHYES 180		
181 STIDVFDFAFOEMALQVNNKGIACGDDELPRITANVPPVYVYGTGEENDPQARNIVASTEG 240			181 STIDVFDFAFOEMALQVNNKGIACGDDELPRITANVPPVYVYGTGEENDPQARNIVASTEG 240		
181 DINDVFDFAFOEMALQVNNKGIACGDDELPRITANVPPVYVYGTGEENDPQARNIVASTEG 240			181 DINDVFDFAFOEMALQVNNKGIACGDDELPRITANVPPVYVYGTGEENDPQARNIVASTEG 240		
241 TTFDVFVNTFYDTFYIPAYGHNHVLNSLAIALACHEEIDSSITKHALKSEFGVARRFN 300			241 TTFDVFVNTFYDTFYIPAYGHNHVLNSLAIALACHEEIDSSITKHALKSEFGVARRFN 300		
241 TAFDVFVNTFYDTFYIPAYGHNHVLNSLAIALACHEEIDSSITKHALKSEFGVARRFN 300			241 TAFDVFVNTFYDTFYIPAYGHNHVLNSLAIALACHEEIDSSITKHALKSEFGVARRFN 300		
301 EKQGDQVLLIDVAHHPTETIKVTLEAARQKYPRDEIYVAFQPHFTFTTQGLDEFASIS 360			301 EKQGDQVLLIDVAHHPTETIKVTLEAARQKYPRDEIYVAFQPHFTFTTQGLDEFASIS 360		
301 ETTIANQVIYDVDAHHPREISATIDTARKKPRHEVAVAFQPHFTFTTQGLDEFASIS 360			301 ETTIANQVIYDVDAHHPREISATIDTARKKPRHEVAVAFQPHFTFTTQGLDEFASIS 360		
361 GADCVYLLIDIGSGARENNKGLTIIDDLCKIKHNAKLEEDDTSVLKADKAVLIFMGAGDI 420			361 GADCVYLLIDIGSGARENNKGLTIIDDLCKIKHNAKLEEDDTSVLKADKAVLIFMGAGDI 420		
361 KADRFELCEITGSIENSGALTIIDDLCKIKHNAKLEEDDTSVLKADKAVLIFMGAGDI 420			361 KADRFELCEITGSIENSGALTIIDDLCKIKHNAKLEEDDTSVLKADKAVLIFMGAGDI 420		
421 OKYMRAY 427			421 OKYMRAY 427		
421 OKYMRAY 427			421 OKYMRAY 427		
RESULT 2			RESULT 2		
ENTRY B70418			ENTRY B70418		
TITLE UDP-N-acetylmutamate-lyase - Aquifex aeolicus			TITLE UDP-N-acetylmutamate-lyase - Aquifex aeolicus		
ORGANISM			ORGANISM		
DATE 08-May-1998			DATE 08-May-1998		
21-Aug-1998			21-Aug-1998		
B70418					

REFERENCE	A70300
#authors	Deckert, G.; Warren, P. V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; AuJay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.
#journal	Nature (1998) 392:353-358
#title	The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
#cross-references	MolMap:9819686
#accession	B70418
#status	preliminary: nucleic acid sequence not shown; translation not shown
##molecule_type	DNA
##residues	31454 ##label AOF
##cross-references	GB:AE000736; NID:g2983763; PID:g2983764; GB:AE000657
##experimental_source	strain VFS
GENETICS	
#gene	murC
CLASSIFICATION	murC
SUMMARY	superfamily UDP-N-acetylmuramate--alanine ligase #length 454 #molecular_weight 50893 #checksum 2346
Query Match	19.9%; Score 620; DB 2; Length 454;
Best Local Similarity	30.8%; Pred. No. 2,698-77;
Matches 139; Conserved	125; Mismatches 164; Indels 24; Gaps 20;
Db	6 IKRHFPGIGIGMSGAOILLLEMKYVSGSDISEKNKTL-LKOKGAKIYGRHPNUG 64
Qy	1 MTHHFVFGIGSGSSSLAQIMHDLGHEVQSGDIENVFTEVALRMKGIRI-LPFGANNIK 59
Db	65 DAOYVYSSAVKPNPEIOEAKRRNIPVIRPGEMLAELKLEGTAVSSGSHQTTTSMI 124
Qy	60 EDWVYVIGKAFASHEITVRAHOLKIDVVSYNDFLQIIDQYTSVAVVGABGKISTTGL 119
Db	125 AEILINAGLEPTVYIGGRKLTGNKLTGSGELVSEADESGSEFLQIPAVAVITNVK 184
Qy	120 SHVA-NODKTSFLIGD----GTGNGLESDVFAFEACEYRHRHLSKYPDAIMTNIDF 173
Db	185 EHLDFYENEFERYKEAFQOFNNSVPFPGFAVNLIDPTLAOLYKKSHERYIYGINSPALV 244
Qy	174 DHPYFDINDVDFAPAFEMAHNYK-GI-IAMGDDEHLRK-IE-ADVPYIYVYGFKSDSDI 229
Db	245 RAKMLYLKEGVEYGEVFKGELGRHLIGA-GIHVNVVALATGVA-LE-LGVSFEVFK 301
Qy	230 YAOINIQLTDGTAVDVIDEFPDHF-LSPQYGDHTVLAALVAISTYELKDYV-N-ITK 286
Db	302 KSLSEFNNAERRLLEKGYKNSPVYDYGHPHPEIKAVINSLRDMPKRLNLVYFOPARY 361
Qy	287 EALTFPGVAKRRNETTI-ANQYIVDDYAHHRPEISATIDTRARKYPRKVEYAAVQPIHF 345
Db	362 SRTYLFEDFVKYKLDKIDLTIVDIPASENNVYGSAAELARK-SGAVFAKDKEEVEEK 420
Qy	346 SRTQAFINFEAESICKRADVFLCEIFGSIRESNGALITDLDIKDIGASFI-N-EDLIN- 402
Db	421 VREHDEGDIVLFLGAGSISKCEFEFLKEYNL 452
Qy	403 VLEQFDNA-VYLFEMGAGDIDKIDNALYDKIGM 433
RESULT	3
ENTRY	C71338
TITLE	probable UDP-N-acetylmuramate--alanine ligase (murC) - syphilis spirochete
ORGANISM	#formal name Treponema pallidum subsp. pallidum #common_name syphilis spirochete
DATE	24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 17-Mar-1999
ACCESSIONS	C71338
REFERENCE	A71250
#authors	Frisser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwinn, M.; Hickey, E.K.; Clayton, R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod, M.P.; Salzberg, S.; Peterson, J.; Kralak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald,



[illegible]



[illegible]

Dd            20 QIHFIIGGAGMGIAEILNEBGTQISGDYADGVYTQ\_RLAAOGAKIITYIGAHEHIEGA 78  
     : ||||| |-----:||:::| : |||| : || : : : :



QY	3	HYHFGIGSGSMSSIAQIMHDLGHEVQSGDIENVFTEVALRNKGKIKI-LPGANNIKED 61
Db	79	SVYVSSAKIKNDNELVYTSKRKRIPYOROMLAELN-RRRHGIANVAGTKTTAMIS 137
QY	62	VIWVGNNAFSSHEIYVAHQIKLDVVSYNDFLQIIDIOTY-SVAATGAGKSTGTGLLS 120
Db	138	MIYQAKIDPFVNGGLVKSAGKNAHIGASRYKILAEADESASFLLQDPVSVVTNNEPD 197
QY	121	HVMNDK-KTIFLLIGD---GTGM-G-LPESDYFPEACEFRHFLSKYKPDYAIMTIDDD 174
Db	198	HMDYEGDFEKKATYVKFLHNLPEFYGLAVWCADDPVLMELVKVGRQVITYGFSQADY 257
QY	175	HPDYK-DINDVFAFGQMAHNVAK-GI-LAMGDDEHLRKIEADY--PIYYGFKSDDI 229
Db	258	RIEDYEQGFQGHYTVICPNNERINVLNLP-GKHNLNLNTAALAAKEGINAEILTA 316
QY	230	YAONIQTIDKGTADY-VYDEFFDHLSPQYGHYTLNLALVIAISTYEKLDPVNIKSA 288
Db	317	LADFGAGRRFDQGEFIRPGRKVLVDGHHTEVGVITKARGSGDKRIYVMEQPH 376
QY	289	LETGCVARRNE-TT-I-AN-QY-IYDDTAHHPREISATIDTARKKYPKHEVAVAFQPH 343
Db	377	RYSTRDLFDDFVOVLNOCVADLMADYVAA 406
QY	344	TESRQAFNLNEFASLSCKADRVFCEIFGS 373
RESULT	8	
ENTRY		CEECAM #type complete
TITLE		UDP-N-acetylmuramate'-alanine ligase (EC 6.3.2.8) - Escherichia coli
ALTERNATE_NAMES		UDP-N-acetylmuramoyl-L-alanine synthetase
ORGANISM		#formal_name Escherichia coli
DATE		31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 02-Jul-1998
ACCESSIONS		J00545; S40601; S65354; C64731
REFERENCE		J00544
#authors		Ikeeda, M.; Wachi, M.; Jung, H.K.; Ishino, F.; Matsunashi, M.
#journal		Nucleic Acids Res. (1990) 18:4014
#title		Nucleotide sequence involving mung and muc in the mra gene cluster region of Escherichia coli.
#cross-references		MUID:90326550
#accession		J00545
#molecule_type		DNA
#residues		1-491 #label IKE
#cross-references		EMBL:X52644; NID:g42053; PID:g42056
REFERENCE		S40531
#authors		Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.; Fujita, N.; Isono, K.; Mizouchi, K.; Nakata, A.
#submission		submitted to the EMBL Data Library, December 1992
#accession		S40601
#molecule_type		DNA
#residues		1-491 #label YUR
#cross-references		EMBL:D10483; NID:g216434; PID:d1001828; PID:g216505
REFERENCE		S65354
#authors		Liger, D.; Masson, A.; Blanot, D.; van Heijenoort, J.; Parquet, C.
#journal		Eur. J. Biochem. (1995) 230:80-87
#title		Over-production, purification and properties of the uridine-diphosphate-N-acetylmuramate-L-alanine ligase from Escherichia coli.
#cross-references		MUID:95324553
#accession		S65354
#status		preliminary
#molecule_type		protein
#residues		1-14 #label LIG
REFERENCE		A64720
#authors		Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Colado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao
#journal		Science (1997) 277:1453-1462

Accession	Gene	Protein	Function	Feature	Summary	Query Match	Best Local Similarity	Matches
#title	The complete genome sequence of <i>Escherichia coli</i> K-12.							
#cross-references	MU01:97428617							
#accession	C64731							
#status	preliminary; nucleic acid sequence not shown; translation not shown							
##molecule-type	DNA							
##residues	1-491 ##label BLAT							
##cross-references	GB:AE000118; GB:U00096; NID:g1786262; PID:g1786279; UWGP:D0091							
##experimental_source	strain K-12, substrain MG1655							
GENETICS								
##gene	murC							
##map_position	2 min							
FUNCTION								
##description	one of the ligases responsible for the synthesis of UDP-N-acetylmuramyl pentapeptide, an intermediate in cell-wall biosynthesis							
##pathway	peptidoglycan biosynthesis							
CLASSIFICATION	##superfamily UDP-N-acetyltransferase-1; alanine ligase							
KEYWORDS	ATP; cell division; cell wall; ligase; P-loop; peptidoglycan biosynthesis							
FEATURE								
126-131	##region nucleotide-binding motif A (P-loop)							
SUMMARY	##length 491 ##molecular_weight 53k26 ##checksum 8296							
Query Match	15.9%; Score 495; DB 1; Length 491;							
Best Local Similarity	31.3%; Pred. No. 8,976-57;							
Matches	123; Conservative 105; Mismatches 138; Indels 27; Gaps 19;							
Db	21 HHFPGJGAGMGGAETLANEGVOISGSDLRPPVVO-QLMNGATITFHHREBNRDA 79							
Oy	3 HHFVGIGSGMSSLAQIMHDLGHEVOGSDIENTVFTEVALRNKGIRI-LPEFGANNKED 61							
Db	80 SVVWVSSAISADNPEIVAAHEARIPVIRRAEMLELM-RFRHGIAIGTHGKTTTAMVS 138							
Oy	62 MVLVIGNNAFASHEIYVAHOLKLDVSYNFDLQIIDQY-SVAVYGARHKSTITGLLS 120							
Db	139 SIYVAGLDPTFFVNGGLVKAAGVHARLGHGRYLIAEDSDASFLHLQPMVAIVTINIEAD 198							
Oy	121 HV-MNGDKRTSFLIGD--GTGM-G-LPESDYFAFECCEYRRHFLSKPYDAIMTINIDF 174							
Db	199 HMDYQGGFENLKGQTFNLFNLFYGRANVCVDPIYRELLPRVNGQOTITGSGSEADV 256							
Oy	175 HPDYFK-DINOVFDFQFQMANHYKK-G-TIAGSDEHRLKEADV--PIYYGGKSDSDI 229							
Db	259 RVEDYQIGPOGHFTLLEKDEKPRKRVETLNAF-GRHNLNNAAVAAVATEEGIDEALIRA 317							
Oy	230 YQNQIITDKSTADIVY-DGEFIDHFLSPYGHYTVLNLAVAIATILEKLDVTINKEA 288							
Db	318 LESFGQGRRRDFLEPEPLFVNGSKSGTAMVDYDGHPTVEVDATIAARAGNPKDLVM 377							
Oy	289 LETEGGVARRN--E-T-TI--AN-QVIVDDYVAHHPRIKSTIDTARKKHYHKEVVA 338							
Db	378 LFGPHRTTRDLYDPRANVLTQVDTLIMEVY 410							
Oy	339 VFQPHFSTRQALNFEFAESLCKADRVFLCEIF 371							
RESULT	9							
ENTRY	S76722							
TITLE	UDP-N-acetylmuramylalanine--D-glutamate ligase (EC 6.3.2.9)							
ALTERNATE_NAMES	murC - <i>Synechocystis</i> sp. (strain PCC 6803)							
ORGANISM	<i>Synechocystis</i> sp.							
KEYWORDS	protein slt1423							
DATE	25-Apr-1997							
ACCESSIONS	21-Aug-1998							
REFERENCE	S76722							
##authors	Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.; Hitosawa, M.; Sugiyama, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,							







```
#status      preliminary; nucleic acid sequence not shown;
```

Conservative 90; Mismatches 156; Indels 23; Gaps 18;



```

Dh 73 GATVATTTAAIPTNPELVEARRRGJRPVLRAPVLAKLAAIGRTLTWCTHCKTTTSM 1322
Qy 61 D-MVYIGNA-FASSHEIYRAHOLKJADVSYNDELQIITDYTSAVTAHKGSTTGL 1185
Dh 133 LIVALOHGDLPRFAVVGELGEATNANHSGDCEFAVEADESDGSLLOYTPHVAITNIE 1922
Qy 119 LSHYM-NODKKTSTLLIG-D-G-TGMGLP-S-DYFAEACEYRRHFLSTYKPDVAIMNID 1722
Dh 193 SDHLDFGVSVEAYVAVPDSFEVERIVPGALVCTDDEFGAALAORATELGIRVLRGVS 2522
Qy 173 FDHDFYEDKINDVEDARQE-MAHNVKKG-IIAMQD---EHL--RKIEADVPYIYGGFKD 2255
Dh 253 GETMAATVSMQOGGVCAGVAHRIKLASLATAOGRVRYKRLVSPGRHMLNMLGALLAAVOI 3122
Qy 226 SDDIYAONI-QITDKGTAFDVYVDGEF-YDH---FLSPQY-GDHTVLAIAVAISYLE 2785
Dh 313 GARPDEVLDLAGLEGVRRRRELVTGCGVAKASRYVEDDYAHNHPTESISATLAARWLEQ 3722
Qy 279 KLDVTINIKEALETGGVKKRRN-ET-TIANQ-VIV-DDYAHNHPREISATIDAPRRKKYPH 3333
Dh 373 GDGRCRWVPEPHLYSTKFAAFEGFALNADEVFVLDDYGA-REOPLGVSAGSAVAEH 4312
Qy 334 KEY---VAVPEPHFSTQALNEFASLCKADVVEFCEJFGSIRENS-GALTIQDILDK 3899
Dh 432 VTVEMRYVPDESAVAQOVAALASPAGDVIYMGAGDVLLLPETILTALRV 4812
Qy 390 IGGA-STFNE-DLIN--VLEOFD-NAVYVLMGAGDIDKLNALYLDKLGKM 4342

```

```

RESULT      15
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references W01D:93171879
#contents      168
#accession     B47691
#molecule_type DNA
#residues      1-494 ##label DAN
#cross-references EMBL:Z15056; NID:g40160; PID:g40162
##note         sequence extracted from NCBI backbone (NCBIN:125659,
                NCBIR:125661)
REFERENCE
#authors
#journal
#title
#cross-references W01D:94118264
#accession     S43864
#type complete
UDP-N-acetylmutamoylalanyl-D-glutamate--2,6-diaminopimelate
ligase (EC 6.3.2.13) mure - Bacillus subtilis
#formal_name Bacillus subtilis
19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change
24-Sep-1998
B47691; S43864; E69662; S23914
B47691
Daniel, R.A.; Errington, J.
J. Gen. Microbiol. (1993) 139:361-370
DNA sequence of the mure-murd region of Bacillus subtilis
168.

```

```

REFERENCE
#authors      S43862
#journal      J. Mol. Biol. (1984) 235:209-220
#title        The Bacillus subtilis spoVD gene encodes a
               mother-cell-specific penicillin-binding protein required
               for spore morphogenesis.
#cross-references EMBL:J25865; NID:g40160; PID:g40162
               sequence extracted from NCBI backbone (NCBIN:125659,
               NBRP:125661)
#accession    S43864
#status       preliminary; nucleic acid sequence not shown;
               translation not shown
##molecule-type DNA
##residues    1-16 ##label DA2
##cross-references EMBL:J25865; NID:g397893; PID:g397896
               the nucleotide sequence was submitted to the EMBL Data
               Library, September 1993
REFERENCE
#authors      Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
A69580

```

Alouin, G.; Azevedo, V.; Belloir, M.G.; Bessières, F.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Bran, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brüsch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;



Daniel, R.A.; Denizot, F.; Devine, K.M.; Duysterhoef, A.;  
 Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;  
 Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,  
 M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Chim,  
 S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grand, G.;  
 Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,  
 C.R.; Henaut, A.; Hilbert, H.; Holstapel, S.; Hosono, S.;  
 Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;  
 Kasahara, Y.; Klaer-Blanchard, M.; Klein, C.; Kobayashi,  
 Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;  
 Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;  
 Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;  
 Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,  
 M.; Moestl, D.; Nakai, S.; Nodack, M.; Noone, D.; O'Reilly,  
 M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro,  
 V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,  
 A.M.; Prescan, E.; Pulic, P.; Purnelle, B.; Rapoport, G.;  
 Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;  
 Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;  
 Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;  
 Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,  
 B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;  
 Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;  
 Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;  
 Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;  
 Wambutt, R.; Wedler, E.; Wedler, H.; Weltzenegger, T.;  
 Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,  
 K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;  
 Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256  
 #title The complete genome sequence of the Gram-positive bacterium  
 Bacillus subtilis.

#cross-references MWID:98044033

#accession E69662 preliminary; nucleic acid sequence not shown;

#status translation not shown

#molecule-type DNA

#residues 1-494 #label KUN

#cross-references GB:29911; GB:AL009126; NID:g2633699; PID:e1185108;

#experimental-source strain 168

# GENETICS

#gene mure

#map-position 133 (degrees)

KEYWORDS ligase: peptidoglycan biosynthesis

SUMMARY #length 494 #molecular-weight 54325 #checksum 6945

Query Match 5.8%; Score 180; DB 2; Length 494;

Best Local Similarity 32.3%; Pred. No. 7.76e-09;

Matches 42; Conservative 22; Mismatches 63; Indels 3; Gaps 3;

DB 251 KYTAA-HISTYGIKNDADYMAKISITAOGTSFDL-VTKGTRHITMSLVGQPNYNTA 308

QY 211 KIEADVPIYYGKSDSDIYAQNIQITDKGTADYVDGEFDHFLSPQYGDHTVNALA 270

DB 309 AVATCTAAGIPPEITTEAVEELGVGRFELVQGOEFPIYDVAHTPDSLENVETCD 368

QY 271 VIAISTLEKLDVTNKEALETFEGVARRNETTIANOV-IVDDYAHHPREISATIDTARK 329

DB 369 MTEGKLEVVV 378

QY 330 KYPHKEVAV 339

Search completed: Wed Nov 24 00:38:32 1999  
 Job time : 118 secs.



**THIS PAGE BLANK (USPTO)**



\*\*\*\*\*  
WIDEVIEW (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Nov 24 00:31:14 1999; Maspar time 13.64 Seconds

Tabular output not generated. 905.767 Million cell updates/sec

Title: >US-09-103-287-2  
Description: (1-437) from US09103287.pep  
Perfect Score: 3121  
Sequence: 1 MTHYHVGIKSGSMSSLAQI.....GDIQKLNAYLDKLGKMAF 437

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 51.864; Variance 106.076; scale 0.489

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	3035	97.2	437	1	MURC_STAUP	0.00e+00
2	1975	63.3	432	1	MURC_BACSU	0.00e+00
3	602	19.3	481	1	MURC_TREPA	8.82e-88
4	581	18.6	468	1	MURC_BORBU	1.07e-83
5	529	16.9	475	1	MURC_HAEIN	1.21e-73
6	495	15.9	491	1	MURC_ECOLI	4.02e-67
7	492	15.8	433	1	MURC_PORCI	1.50e-66
8	477	15.3	483	1	MURC_BUCAP	1.09e-63
9	470	15.1	503	1	MURC_STINY3	2.33e-62
10	440	14.1	453	1	MPL_HAEIN	1.10e-56
11	424	13.6	457	1	MPL_ECOLI	1.12e-53
12	422	13.5	449	1	MURC_HELPI	2.66e-53
13	349	11.2	494	1	MURC_MYCTU	8.50e-40
14	337	10.8	280	1	MURC_STYND7	1.29e-37
15	194	6.2	456	1	MURC_ENTFA	4.90e-13
16	180	5.8	494	1	MURC_BACSU	7.63e-11
17	177	5.7	457	1	MURC_HAEIN	2.21e-10
18	164	5.3	460	1	MURD_ENTHR	2.05e-08
19	158	5.1	505	1	MURC_STINY3	1.58e-07
20	142	4.5	508	1	MURC_BORBU	3.07e-05
21	128	4.1	449	1	MURD_STAUP	2.44e-03
22	128	4.1	457	1	MURC_BACSU	2.44e-03
23	121	3.9	464	1	MURF_BORBU	1.96e-02

24	118	3.8	451	1	MURD_BACSU	4.70e-02
25	116	3.7	437	1	MURD_HAEIN	8.33e-02
26	113	3.6	320	1	OCB_RHIME	1.94e-01
27	111	3.6	432	1	Y672_METJA	3.39e-01
28	109	3.5	182	1	APT_PSEAE	5.86e-01
29	109	3.5	452	1	MURF_ECOLI	5.86e-01
30	105	3.4	227	1	Y295_METJA	1.72e+00
31	105	3.4	234	1	TRAR_AGRVI	1.72e+00
32	105	3.4	378	1	PGK_CONNG	1.72e+00
33	105	3.4	395	1	ENO_ALIMI	1.72e+00
34	106	3.4	416	1	CINA_BACSU	1.32e+00
35	105	3.4	758	1	YK04_YEAST	1.72e+00
36	105	3.4	806	1	SUBV_BACSU	1.72e+00
37	103	3.3	433	1	ENOA_CHICK	2.90e+00
38	103	3.3	476	1	THB2_HUMAN	2.90e+00
39	102	3.3	477	1	OPRM_PSEAE	3.76e+00
40	102	3.3	750	1	PBPX_STRPN	3.76e+00
41	103	3.3	764	1	DPO3_SULSO	2.90e+00
42	101	3.2	226	1	Y703_METJA	4.86e+00
43	100	3.2	234	1	TRAR_AGR15	6.28e+00
44	101	3.2	433	1	ENOA_ANAPL	4.86e+00
45	100	3.2	1013	1	DPOL_NPVLD	6.28e+00

## ALIGNMENTS

RESULT	ID	1	STANDARD	PRT	437 AA.
AC	031211				
DT	15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)				
DT	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)				
DE	UDP-N-ACETYLURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-				
DE	ACETYLURAMATE--L-ALANINE SYNTHETASE).				
GN	MURC.				
OS	STAPHYLOCOCCUS AUREUS.				
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;				
OC	STAPHYLOCOCCUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	LOWE A.M., DERESEWICZ R.L.;				
RL	SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.				
CC	-1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).				
CC	-1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLURAMATE + L-ALANINE -				
CC	ADP + ORTHOPHOSPHATE + UDP-N-ACETYLURAMATE--L-ALANINE.				
CC	-1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.				
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).				
CC	-1- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
DR	EMBL: AF034076; G2642659;				
KW	PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;				
KW	ATP-BINDING.				
FT	NP_BIND 108 114 ATP (POTENTIAL);				
SO	SEQUENCE 437 AA; 49176 MW; 64068582 CRC32;				
Query Match					
Best Local Similarity 97.2%; Score 3035; DB 1; Length 437;					
Matches 426; Conservative 5; Mismatches 6; Indels 0; Gaps 0;					
DB	1	MTHYHVGIKSGSMSSLAQIMHDLGHEVSGDIENVFVFLRNKGITLPDANNINE 60			
QY	1	MTHYHVGIKSGSMSSLAQIMHDLGHEVSGDIENVFVFLRNKGITLPDANNINE 60			
DB	61	DNVVIOGNFASSHKRAHQKLDVVSYNDFLGOIIOOYVSVAVTGAHGTSTGLS 120			



QY 61 DMVYIQAFASSHEEIVRAHQKLDVVSYNDELGOIIDQTSVAVGANGKSTTGLS 120  
 Db 121 HVMGDKKTSFLIGDGTGMGLPESDYFAFEACEYRRHFLSKYKPDYAIMTINIDFHPDYFK 180  
 QY 121 HVMGDKKTSFLIGDGTGMGLPESDYFAFEACEYRRHFLSKYKPDYAIMTINIDFHPDYFK 180  
 Db 181 DINDVFAFOEMAHNVKGGIIMAGDDEHLRKIEADVPIYYGFKSDSDIYAQNIDITDKG 240  
 QY 181 DINDVFAFOEMAHNVKGGIIMAGDDEHLRKIEADVPIYYGFKSDSDIYAQNIDITDKG 240  
 Db 241 TAPDVYVGEFEYDHFSLPOYGDHVTNLAVATSLYLEKLDVTNKLELTFEGVKRRFN 300  
 QY 241 TAPDVYVGEFEYDHFSLPOYGDHVTNLAVATSLYLEKLDVTNKLELTFEGVKRRFN 300  
 Db 301 ETTIANQYIVDYAHHPREISATIEETARKKYPRHKEVAVAFOPHTFSRTOAFLEFEASLS 360  
 QY 301 ETTIANQYIVDYAHHPREISATIDTARKKYPRHKEVAVAFOPHTFSRTOAFLEFEASLS 360  
 Db 361 KADRYFLCEIFGSIRESNGALTIODLIDKIGASLINESDINLEQPDNAVVLFGAGDI 420  
 QY 361 KADRYFLCEIFGSIRESNGALTIODLIDKIGASLINESDINLEQPDNAVVLFGAGDI 420  
 Db 421 OKLONAYLDKLGKMAF 437  
 QY 421 OKLONAYLDKLGKMAF 437  
 RESULT 2  
 ID MURC\_BACSU STANDARD: PRT: 432 AA.  
 AC P40778;  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-  
 DE ACTYLMURANOYL-L-ALANINE SYNTHETASE).  
 GN MURC.  
 OS BACILLUS SUBTILIS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;  
 CC BACILLUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE; 98048467.  
 RA LAFIDUS A., GALLERON N., SOROKIN A., EHRLICH S.D.;  
 RT "Sequencing and functional annotation of the Bacillus subtilis genes  
 in the 200 kb rnb-dnaB region."  
 RL MICROBIOLOGY 143:3431-3441(1997).  
 RN [2]  
 RP SEQUENCE OF 85-432 FROM N.A.  
 RC STRAIN-168 / MARBURG;  
 RX MEDLINE; 96310371.  
 RA VARON D., BRODY M.S., PRICE C.W.;  
 RT "Bacillus subtilis operon under the dual control of the general  
 stress transcription factor sigma B and the sporulation transcription  
 factor sigma H."  
 RL MOL. MICROBIOL. 20:339-350(1996).  
 CC -1- FUNCTION: CELL WALL FORMATION.  
 CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL-L-ALANINE -  
 CC -1- ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.  
 CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE MURCEDEE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF008220: G2293216;  
 CC EMBL: U31845: G556014;  
 CC SUBTILIST; BG10973; MURC.

KW PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;  
 KM ATP-BINDING.  
 FT NP\_BIND 108 114 ATP (POTENTIAL).  
 SO SEQUENCE 432/AA; 48364 MW; 163166CB CRC32.  
 Query Match 63.3%; Score 1975; DB 1; Length 432;  
 Best Local Similarity 63.9%; Pred. No. 0.00e+00;  
 Matches 273; Conservative 60; Mismatches 94; Indels 0; Gaps 0;  
 Db 1 MTHYFVGIKSGSPILAOILHNDGYVVGSDIEKFTFTQALAEKRNITLPSAEINIKP 60  
 QY 1 MTHYFVGIKSGSPILAOILHNDGYVVGSDIEKFTFTQALAEKRNITLPSAEINIKP 60  
 Db 61 GMTYIAGNAPFDTPEIEKAMSECIPIRYHKEFLGDYMKFTSVAVTGANGKSTTGLA 120  
 QY 61 DMVYIQAFASSHEEIVRAHQKLDVVSYNDELGOIIDQTSVAVGANGKSTTGLS 120  
 Db 121 HVMGDKKTSFLIGDGTGMGLPESDYFAFEACEYRRHFLSKYKPDYAIMTINIDFHPDYFK 180  
 QY 121 HVMGDKKTSFLIGDGTGMGLPESDYFAFEACEYRRHFLSKYKPDYAIMTINIDFHPDYFK 180  
 Db 181 DINDVFAFOEMAHNVKGGIIMAGDDEHLRKIEADVPIYYGFKSDSDIYAQNIDITDKG 240  
 QY 181 DINDVFAFOEMAHNVKGGIIMAGDDEHLRKIEADVPIYYGFKSDSDIYAQNIDITDKG 240  
 Db 241 TAPDVYVGEFEYDHFSLPOYGDHVTNLAVATSLYLEKLDVTNKLELTFEGVKRRFN 300  
 QY 241 TAPDVYVGEFEYDHFSLPOYGDHVTNLAVATSLYLEKLDVTNKLELTFEGVKRRFN 300  
 Db 301 ETTIANQYIVDYAHHPREISATIEETARKKYPRHKEVAVAFOPHTFSRTOAFLEFEASLS 360  
 QY 301 ETTIANQYIVDYAHHPREISATIDTARKKYPRHKEVAVAFOPHTFSRTOAFLEFEASLS 360  
 Db 361 KADRYFLCEIFGSIRESNGALTIODLIDKIGASLINESDINLEQPDNAVVLFGAGDI 420  
 QY 361 KADRYFLCEIFGSIRESNGALTIODLIDKIGASLINESDINLEQPDNAVVLFGAGDI 420  
 Db 421 OKYMRAY 427  
 QY 421 OKLONAY 427  
 RESULT 3  
 ID MURC\_TREPA STANDARD: PRT: 481 AA.  
 AC O83361;  
 DT 15-DEC-1998 (REL. 37, CREATED)  
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
 DE UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-  
 DE ACTYLMURANOYL-L-ALANINE SYNTHETASE).  
 GN MURC OR TP0341.  
 OS TREPONEMA PALLIDUM.  
 OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; TREPONEMA.  
 CC [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NICHOLS;  
 RX MEDLINE; 98332770.  
 RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,  
 RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,  
 RA SODERREN E., HARDHAM J.M., MCLEOD M.P., SALBERG S., PETERSON J.,  
 RA KHALAF H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTERBACK T.,  
 RA MC DONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,  
 RA HATCH B., HORST K., ROBERTS K., MATTHEY L., WEIDMAN J., SMITH H.O.,  
 RA VENTER J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT spirochete."  
 RL SCIENCE 281:375-388(1998).  
 CC -1- FUNCTION: CELL WALL FORMATION.  
 CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL-L-ALANINE -  
 CC -1- ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.  
 CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE MURCEDEE FAMILY.



[illegible]

Rt	bungdorferi . "
RL	NATURE 390:580-586(1997).
CC	-1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).
CC	-1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL + L-ALANINE =
CC	ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.
CC	-1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC	-1- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
CC	-----
CC	This SWISS-Prot entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see <a href="http://www.isd-sib.ch/announce/">http://www.isd-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).
CC	-----
DR	EMBL; AE001180; G2688761; -
DR	TIGR; B08017; -
KM	PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;
KM	ATP-BINDING.
FT	NP_BIND 121 127 ATP (POTENTIAL);
SQ	SEQUENCE 468 AA; 53480 MM; 09511A39 CRC32;
	-----
	Query Match 18.6%; Score 581; DB 1; Length 468;
	Best Local Similarity 32.3%; Pred. No. 1,07e-83;
	Matches 149; Conservative 125; Mismatches 144; Indels 43; Gaps 33;
D6	13 FVGINGSGAGSLACFLGSKGYCEGVSDYDKRYTDELISNNKISYDNIYEFSLKOLDS 72
OY	::   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
	6 FVGINGSGAGSLAQIMHDLGHEVGSDIENYFTEVALRN-K-GI-KLPFGANNIKED 61
D6	73 FDLIIVSSAYNKDGLOVLLEAKELNIPILSYPEALGEISRKYYSIGIAGSHGKTITTAFL 132
OY	::::  : ::::  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
	62 M-VVIOGNFA-SSEHEIYRAHQKLIDVYSNDIFGOIIDOYTSVAVGAHGKSTTGLL 119
D6	133 GYLEFKKLGINPVYIVGSSVKDFKNDSALAGISINITEVECEKKRFNFSPMLLTIND 192
OY	120 SHVMNG-DKRTSFLLGDGT-G-M---GLPE-SDYAFACEVRHRFLSKPPYAIAITNTID 172
D6	193 YEHVOFFKRYEALAEAPLOYINNLKKGILLINSDDNMLLKRPDIRKDIFSFGSD 252
OY	:::  : ::::  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
	173 PDHPPIFYFDINVPFAFOEMAHNVK-GI-TMGDEHL-R-K--IEA-DVIYIYGGFD 225
D6	253 LSDFOISINAVNSEYFCFS-FL-GLLNVELKTVLF-HNVLSAALLALNPLESNGKS 308
OY	226 SDDIYAQNQIQIDKTAFPDYVDGFHYHFLSPQGDHTVLN-ALAVALIS-YLE--K- 279
D6	309 IFDEDAIKRIAKANTSGIKRREYVYKEENGVIYMDYAHHPREIKNTLFGINKFYKNKRI 368
OY	:  : ::  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
	280 I-DVTN-IKEALETFEGVGRRENETTIANOVY-VDDYAHHPREIASITIDTARKKYPKHEV 336
D6	369 IJDDMPHFTRKEPFADVEVLSAADILLINIYLSNRENNPDLSVKLRLNKTIK 428
OY	:     :     :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
	337 VAVFDPHTFSKIQAFLNEFABSLCADRVFLDELIGSIREN--SALTIIQD-L-IDKIGG 392
D6	429 NTYFEKDVASINFIRKSLISGDLTITTGAGNNFIILD-FL 468
OY	:  : ::  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
	393 AS-FINE--DLINVLEQFDMAVYLF--WGAGDIQLQNAYL 428
	-----
RESULT	5
ID	MURC_HAEIN STANDARD; PRT; 475 AA.
AC	P45066;
DT	01-NOV-1995 (REL. 32, CREATED)
DT	01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE	UDP-N-ACETYLMURAMATE--ALANINE LIASE (EC 6.3.2.8) (UDP-N-
DE	ACETYLMURAMOYL-L-ALANINE SYNTHETASE).
GN	MURC OR H11139.
OS	HAEMOPHILUS INFLUENZAE.
OC	BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURILLACEAE;
OC	HAEMOPHILUS.



```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20:
RX MEDLINE; 95350630.
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA KRELAJAVE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA KENNEDY K., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
RA MEDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,
RA UTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDER D.M., BRANDON R.C.,
RA FINE L.D., FITCHMAN J.L., FUHRMANN J.L., GEOGAGEN N.S.M.,
RA GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA VENTER J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL SCIENCE 269:496-512(1995).
CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLURAMONYL + L-ALANINE =
CC -1- ADP + ORTHOPHOSPHATE + UDP-N-ACETYLURAMONYL-L-ALANINE.
CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32794; G1574695; -.
DR TIGR: H11139; -.
KW PERITOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;
KW ATP-BINDING.
FT NP_BIND 125 131 ATP (POTENTIAL).
SQ SEQUENCE 475 AA; 51994 MW; DE40C270 CRC32;

Query Match 16.9%; Score 529; DB 1; Length 475;
Best Local Similarity 30.5%; Pred. No. 1,216,73;
Matches 119; Conservative 104; Mismatches 145; Indels 22; Gaps 19;

Db 20 QIHFGLGAGMSGIAEILNEGYSISDIADGVVTO-RLAQAQAKIYIGAEHEHIGA 78
QY 3 HYHFVGLGSGMSLAQIMHDLGHEVGSDEIENVYFTEALRNKGIK-LFGANNIMED 61
Db 79 SVYVYSSAIKDNPELVTSKQRIYVIOAQLAII-M-RFRHGLAVAGTGGKTTTAMIS 137
QY 62 MVVIOGNAFASHEIVRAHOLKLDVVSYNDFLGQIIDQYT-SVAVTGAHGKSTTGLLS 120
Db 138 MIYQAQKADPFVNGVLKVSAGKNAHLGASRYLLIAEDSDASFHLDPMSVYVNMEDP 197
QY 121 HYVWMDK-KTFELID---GTGM-G-LPESDYFAFEACEYRRHFLSYFDYALIMNIDFD 174
Db 198 HMDYEGDFEKKATYVKEFLNLPFYGLAVMCADDPVLMELVPKGVROVITYGSEQADY 257
QY 175 HPDYFK-DINVDFAFQMAHVVK-GI-IAWGDEHLRKIEADV--PIYYGFGDSDDI 229
Db 258 RIEDYEGFGFGHIVTGPNNERIVLNV-P-GKHNALNAPALAAVAKEEGIANALIEA 316
QY 230 YAQMIIQIDKSTADPV-VYDEFDYHFLSPQGDHTVINALAVIAYISLEKIDVTNKEA 288
Db 317 LADGAGARRDOLGEFTRPNGKVLVDYDGHHPTEGVTVTKAAEGSGDKRIYVIRPQH 376
QY 289 LETGGVARRRNE-IT-I-AN-QV-IVDYAHHPREISATIDTARKKTPHREKVAVFPQH 343
Db 377 RYSTRDLDFDEVYVLSQVDALIMDYAA 406
QY 344 TFSRTQALNFEASLCKADRVFLCEIRGS 373

RESULT 6
ID MURC_ECOLI STANDARD; PRT; 491 AA.

AC P17952; 007099;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE UDP-N-ACETYLURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-
DE ACETYLURAMONYL-L-ALANINE SYNTHETASE).
GN MURC.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 90326550.
RA IKEDA M., NACHI M., JUNG H.K., ISHINO F., MATSUHASHI M.;
RT "Nucleotide sequence involving murG and murC in the mra gene cluster
RT region of Escherichia coli.";
RL NUCLEIC ACIDS RES. 18:4014-4014(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 92334977.
RA YODA T., MORI H., NAGAI H., NAGATA T., ISHIMURA A., FUJITA N.,
RA ISONO K., MIZOBUCHI K., NAKATA A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.";
RL NUCLEIC ACIDS RES. 20:3305-3308(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., ROSE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL SCIENCE 277:1453-1474(1997).
RN [4]
RP MUTANT MURC3, SEQUENCE FROM N.A.
RC STRAIN-CGSC 5988;
RX MEDLINE; 97309380.
RA EVELAND S.S., POMPLIANO D.L., ANDERSON M.S.;
RT "Conditionally lethal Escherichia coli murein mutants contain point
RT defects that map to regions conserved among murein and folyl
RT poly-gamma-glutamate ligases: identification of a ligase
RT superfamily.";
RL BIOCHEMISTRY 36:6223-6229(1997).
RN [5]
RP SEQUENCE OF 1-14, AND CHARACTERIZATION.
RX MEDLINE; 95324553.
RA LIGER D., MASSON A., BLANOT D., VAN HEIJENOORT J., PARQUET C.;
RT "Over-production, purification and properties of the
RT uridine-diphosphate-N-acetylmuramate-L-alanine ligase from
RT Escherichia coli.";
RL EUR. J. BIOCHEM. 230:80-87(1995).
CC -1- FUNCTION: CELL WALL FORMATION.
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLURAMONYL + L-ALANINE =
CC -1- ADP + ORTHOPHOSPHATE + UDP-N-ACETYLURAMONYL-L-ALANINE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52644; G42056; -.
DR EMBL: X55034; G40859; -.
DR EMBL: D10483; G216505; -.

```



DR EMBL: AE000118; G1786279; -  
 DR EMBL: U67892; G2177094; -  
 DR PIR: J00545; CECAM.  
 DR PIR: S40601; S40601.  
 DR ECOGENE: EGI0619; MURC.  
 DR PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;  
 KW ATP-BINDING.  
 FT NP\_BIND 126 132 ATP (POTENTIAL).  
 FT MUTAGEN 344 344 G->D: IN MURC3.  
 SQ SEQUENCE 491 AA; 53626 MW; C68ED87D CRC32;

Query Match 15.98; Score 495; DB 1; Length 491;  
 Best Local Similarity 31.38; Pred. No. 4,02e-67;  
 Matches 123; Conservative 105; Mismatches 138; Indels 27; Gaps 19;

Db 21 HIFHVGIGSGMGIAEVLNNEGVOISGDLAPNPVQ-OLMUGATITYNHREPNRDA 79  
 QY 3 HIFHVGIGSGMGSLAQMIDHLEVOGSDIENTVFTEVALRNKGIKI-LPFGANNIKED 61  
 Db 80 SYVVVSSASADNPETVAHEARIPVIRRAEMLAELM-REFHGIATGHTGTTTTAMVS 138  
 QY 62 MVYIQGNAPFSSHEIYRAHQKLDVSYNDLFGQIIDQYT-SVAVTGAGKSTGGLS 120  
 Db 139 STYAEAGLDFEVNGGLVKAQVHARLGHRYLIAEDSDASFLLQPMVAIVTNEAD 198  
 QY 121 HV-MNGDKTSFLIGD---GTGM-G-LPESDYFAFACERYRRHFLSKYKPDYAIINTNIDF 174  
 Db 199 HMTYOGDEPNLKQITLNFHNPFGRAVMCVDDPIRELLPRVGQTTTGFSEADY 258  
 QY 175 HPDYFK-DINDVDARQEMAHNVK-K-G-IIAMGDDEHLRKIEADY--PIYYSFKSDDI 229  
 Db 259 RVDYDQIQGQGFHTLRQDEKPRVLTNP-GRHNALNAAVAATEGIDEALTRA 317  
 QY 230 YANQIQITQKGFADYVY-DGEFYDHLSPQYGDHTVNLAVIAISYLEKLDVTNKEA 288  
 Db 318 LESFOGTGRFDFLGEFFLEPPVNGKSGTAMLVDDYGHHTPEVDATTKARAGMPDKNLV 377  
 QY 289 LEFFGGVKKRFN---E-T--TI--AN-QVIYDDYAHHPREISATIDTARKKYPHKEVVA 338  
 Db 378 LEPHRTTRDLYDDPFAVLTOVDITLMEVY 410  
 QY 339 VFQPHFESRTQAFLEPFAESLCKRADRVFLCEIF 371

RESULT 7  
 ID MURC\_PORGI STANDARD; PRT; 433 AA.  
 AC 051831;  
 DT 15-JUL-1998 (REL. 36, CREATED)  
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE UDP-N-ACETYLURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-  
 DE ACETYLMURANOYL-L-ALANINE SYNTHETASE).  
 GN MURC.  
 OS PORPHYROMONAS GINGIVALIS (BACTEROIDES GINGIVALIS).  
 OC BACTERIA; CYTOPHAGALES; BACTEROIDACEAE; PORPHYROMONAS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-381;  
 RX MEDLINE: 96118684.  
 RA ANSAI T.; YAMASHITA Y., AMANO S., SHIBATA Y., WACHI M., NAGAI K.,  
 RA TAKEHARA T.;  
 RT "A murc gene in Porphyromonas gingivalis 381."  
 RL MICROBIOLOGY 141:2047-2052(1995).  
 RU SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURANOYL + L-ALANINE -  
 CC ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURANOYL-L-ALANINE -  
 CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).

CC -1- SIMILARITY: BELONGS TO THE MURCER FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: D84504; D1025270; -  
 DR PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;  
 KW ATP-BINDING.  
 FT NP\_BIND 114 120 ATP (POTENTIAL).  
 SQ SEQUENCE 433 AA; 48309 MW; 38C4714B CRC32;

Query Match 15.88; Score 492; DB 1; Length 433;  
 Best Local Similarity 31.78; Pred. No. 1,50e-66;  
 Matches 126; Conservative 96; Mismatches 149; Indels 26; Gaps 19;

Db 1 MKRHYFVIGIGSMASIAHYEAKGNVCGYDLPSPID-OLIKEGIEV-HRSDDLMI 58  
 QY 1 MTHYFVIGIGSGMSLQIMHDLGHEVOGSDIENTVFTEVALRNKGIKILPFGAN-N-I 58  
 Db 59 PKAFESPTSLIYVTPAVPADHSELTYFRSNGYRVKRAEVLGEITLMERALCVAAGTK 118  
 QY 59 -KE-----DMVYIQGNAPFSSHEIYRAHQKLDVSYNDLFGQIIDQYTSVAVTGAKK 112  
 Db 119 TTTSTLLAHLKQSHVDCAFLGGSNNYQSNLLSDKSDLVVEADEDFRSFHLKPRM 178  
 QY 113 TSTTGLSHVMN-GDKK-TSFL--IGDG--SDYFAFACERYRRHFLSKYKPDY 165  
 Db 179 AITTSADPHMDYGAENYVRDSFEHETSLIOSGALVLYKAPVPR-LGSPVSLFST 237  
 QY 166 AITNTIDFHPD-YEKDIN--DVDFAPQEMAHNVKKGIIAMGDDEHLRKIEADVPITYIG 222  
 Db 238 SDDRQADYFASDIMIDRGLEFTMHPGQGLEVELEVGVPRIN-VENAAMAAIAHLNGV 296  
 QY 223 FKNSD-DIYAQNQITDKTAPD-VYVDEFDYDHLSPQYGDHTVNLAVIAISYLEK 280  
 Db 297 TVEELRSGIASFKSHREKRVKLTREVVLLIDYAHHPVELDAIRSVREIYSGKHIMCI 356  
 QY 281 DVYNIKEALETFEGVARRRNETTIANQVI-VDDYAHHPREISATIDTARKKYPHKEVAV 339  
 Db 357 FQPHLSRTADYQDFAKSLSMIDQVLLDIYPA-RE 392  
 QY 340 FQPHFESRTQAFLEPFAESLCKRADRVFLCEIFGSI 376

RESULT 8  
 ID MURC\_BUCAP STANDARD; PRT; 483 AA.  
 AC 051926;  
 DT 15-DEC-1998 (REL. 37, CREATED)  
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-  
 DE ACETYLMURANOYL-L-ALANINE SYNTHETASE).  
 GN MURC.  
 OS BUCHNERA APHIDICOLA.  
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; BUCHNERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE: 96087557.  
 RA BAUMANN L., BAUMANN P.;  
 RT "Characterization of ftsZ, the cell division gene of Buchnera  
 RT aphidicola (endosymbiont of aphids) and detection of the product."  
 RL CURR. MICROBIOL. 36:85-89(1998).  
 CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURANOYL + L-ALANINE -  
 CC ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURANOYL-L-ALANINE -  
 CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE MURCER FAMILY.







RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,  
RA FINE L.D., FRITZMAN J.L., FUHRMAN J.L., GORGAGEN N.S.M.,  
RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,  
RA VENTER J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
influenzae Rd.";  
RT SCIENCE 269:496-512(1995).  
CC -I- FUNCTION: INVOLVED IN CELL WALL FORMATION. REUTILIZES THE INTACT  
CC TRIPEPTIDE L-ALANYL-GAMMA-D-GLUTAMYL-MESO-DIAMINOPIMELATE.  
CC -I- SIMILARITY: BELONGS TO THE MURDEE FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: U32698; G1573076; -  
DR TIGR: H10121; -  
KW PEPTIDOGLYCAN SYNTHESIS; CELL DIVISION; CELL WALL; LIGASE;  
KW ATP-BINDING.  
FT NP\_BIND 111 117 ATP (POTENTIAL).  
SQ SEQUENCE 453 AA; 50074 MW; 72A96638 CRC32;  
  
Query Match 14.1%; Score 440; DB 1; Length 453;  
Best Local Similarity 29.7%; Pred. No. 1,10e-56;  
Matches 116; Conservative 92; Mismatches 156; Indels 26; Gaps 21;  
  
Db 1 MKHHLIGCTFMGVMAMAKOMGVHTGSDTNYVPMSFLEEGEIEIENYDVAOLQ 60  
Oy 1 MHHYHEVGKSGMSLAQIMHDLGHEVGSDIENYVFEVALRNKGIILP-FGANNNK 59  
Db 61 PARDWIV-GNAKKGNPCVEYEVLENLKYTSGPOMLHLLRDRV-LAVSGTGGKTTT 118  
Oy 60 -E-DWVVGDMFASHEHIEIVRAHQLKLDVSYNDPL-QQII-DOYTSVAVGAHGKIST 115  
Db 119 TSMLEFVWVONGKSGELGIGTAGNGISARLGDSPYFIIEDEYTAEPDRKRVHN 178  
Oy 116 TGLLSVYVNGDK-KTSFLIGD--GT-GMG--LPESDYFAEA--CE--Y-RR-HFLSK 162  
Db 179 PRTLYNNISFDHADFDLKAIOFHHMIFRTIPASGLVLSASPOSARETLALGWSQ 238  
Oy 163 PVALMTNIDPFHPDYFKINDVFDFAQEMAHNVK-KGIANGDDHLKIEADVPYIYY 221  
Db 239 QOFLGKDNEMFAE-RITNDASHFAVHGEKVAEYKWNVYGOHNMHMLAIAAHHTGV 297  
Oy 222 G-FKSDSDIYAQNIQTIDKGTAFDYVDGEFYDHEFLSPQGDHTVLAALVAIAYLEKL 280  
Db 298 ALEDACKAIGSVNKKRLEVGSEVNSITVYDDFAHHPAILATLALADKVGGRILA 357  
Oy 281 DVTKIKAELETGKVKRRNETTIAQVIV-DDYAHHPREISITIDTARKKIPKHE-VVA 338  
Db 358 VLEPSNTMKMGVHKDDIAPALGRADAVEM 387  
Oy 339 VFQPHFFS-RTQAFLENEFAESLCKADRVFL 367  
  
RESULT 11  
ID MPL\_ECOLI STANDARD; PRT; 457 AA.  
AC P37773; P76804;  
DT 01-OCT-1994 (REL. 30. CREATED)  
DT 15-JUL-1998 (REL. 36. LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37. LAST ANNOTATION UPDATE)  
DE UDP-N-ACETYLIMURAMATE:L-ALANYL-GAMMA-D-GLUTAMYL-MESO-DIAMINOPIMELATE  
DE LIGASE (EC 6.3.2.-).  
GN MPL.  
OS ESCHERICHIA COLI.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
OC ESCHERICHIA.  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;  
RX MEDLINE; 95334362.  
RA BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,  
RA BLATTNER F.R.;  
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
RT region from 92.8 through 100 minutes."  
RT NUCLEIC ACIDS RES. 23:2105-2119(1995).  
RN [12]  
RP SEQUENCE OF 1-64 FROM N.A.  
RX MEDLINE; 88335617.  
RA HAMILTON W.D.O., HARRISON D.A., DYER T.A.;  
RT "Sequence of the Escherichia coli fructose-1,6-bisphosphatase gene."  
RT NUCLEIC ACIDS RES. 16:8707-8707(1988).  
RN [13]  
RP IDENTIFICATION.  
RX MEDLINE; 95075659.  
RA BORODOVSKY M., RUDD K.E., KOONIN E.V.;  
RT "Intrinsic and extrinsic approaches for detecting genes in a  
RT bacterial genome."  
RL NUCLEIC ACIDS RES. 22:4756-4767(1994).  
RN [14]  
RP CHARACTERIZATION.  
RX MEDLINE; 96404780.  
RA MENGIN-LECREUX D., VAN HEIJENOORT J., PARK J.T.;  
RT "Identification of the mpl gene encoding UDP-N-acetylmuramate:  
RT L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase in Escherichia  
RT coli and its role in recycling of cell wall peptidoglycan."  
RL J. BACTERIOL. 178:5347-5352(1996).  
CC -I- FUNCTION: INVOLVED IN CELL WALL FORMATION. REUTILIZES THE INTACT  
CC TRIPEPTIDE L-ALANYL-GAMMA-D-GLUTAMYL-MESO-DIAMINOPIMELATE.  
CC -I- SIMILARITY: BELONGS TO THE MURDEE FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: U14003; G537075; -  
DR EMBL: AE000494; G1790680; -  
DR EMBL: X12545; G1790680; -  
DR ECOGENE: EGI2440; MPL.  
DR HSSP: P14900; IUDG.  
KW PEPTIDOGLYCAN SYNTHESIS; CELL DIVISION; CELL WALL; LIGASE;  
KW ATP-BINDING.  
FT NP\_BIND 110 116 ATP (POTENTIAL).  
SQ SEQUENCE 457 AA; 49874 MW; A2BF8001 CRC32;  
  
Query Match 13.6%; Score 424; DB 1; Length 457;  
Best Local Similarity 29.4%; Pred. No. 1.12e-53;  
Matches 113; Conservative 102; Mismatches 146; Indels 24; Gaps 18;  
  
Db 4 HILGICGTFMGVLAMLARQLGHEVYSDANYRPPMSFLEEGEIEIIOGYDASOLEPPD 63  
Oy 5 HEVGLKSGMSLAQIMHDLGHEVGSDIENYVFEVALRNKGIILP-FGANNNK-E-D 61  
Db 64 LVIT-GNAFTRGNPCVEALEKNIPIYMSGPOMLHDLVLRDRAVLAAGTGGTTGAT 122  
Oy 62 NVYIOGNFASSHIEIVRAHQLKLDVSYNDPLQIIDQYTSV-AVTAHGKISTTGLIS 120  
Db 123 WLEBOGKRPGVIVIGVNGNEVSAHLDESDFYEADEYDCAFDKSKFVHYCPRLI 182  
Oy 121 HVMNG-DKRTSLID--GT-GMG--LPESDYFAEA--CEY--RR-HFLSKYRPAI 167  
Db 183 LNNLEFDHADIPDLDKAIQKOPHHLVRIYVPGGRTIPENDINLEKOTNA-MGCWSEQELV 241  
Oy 168 MTNIDPDHDPYKINDVFDFAQEMAHNVK-KGIANGDDE-HLKIRADVPYIYYGKD 225  
Db 242 GEOGHQAQKLTITDASEWEVILLDEKVEYKWSLVGEHNMHNGMLAIAAHVGAAPADA 301  
Oy 226 SDDIYAQNIQTIDKGTARDVYVDGEFYDHEFLSPQGDHTVLAALVAIAYLEKLDTVNI 285



```

QY      123 -NM-GDKTSLFLIGDGTGGLPESDYFAFEACEYRRHFLSYKPDYA-INTINDPHDPVE 179
Db      195 GDDLEFFFAFYEDLDHAOKRYI-YKEDFLKNYSKN-AI--VL-EKKDIY--NIQYL 246
QY      180 -KDINDVFAPAFEMAHNNKKGIIANGDEDEHLKIKIBADVPYYGYGRDSODIAONIQ-TT 237
Db      247 KCGEPTSPFELDL-GAFLWGLGEENNTNA-SLAILSALDELHEIEIRNNLNFKRGKI 304
QY      238 DKGTAFDVYDGEFDHLSFPQGHVYLNALAIVASTLELDVTYNIKEALETTGGVKR 297
Db      305 REDILOKNALIIDDYAAHHPEISATLSKSARIYANMLNTOEKIIVYQWAHKYSRLMDNLE 364
QY      298 RNETTIANOVYVDYAYHHPREISATIDPAR--KYPHK-E-VVAVFQHTSSRFQAFLN 353
Db      365 EFKKCESEHCRLIIPYISA 385
QY      354 EFAESICKRA-DRVFLCEIFGS 373

RESULT   13
ID        MURC_MCTU          STANDARD:          PRT:    494 AA.
AC         006225;
DT       15-JUL-1998 (REL. 36, CREATED)
DT       15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT       15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE       UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-
DE       ACETYLMURAMOYL-L-ALANINE SYNTHETASE).
MURC OR MUCYZ270.16.
OS       MYCOBACTERIUM TUBERCULOSIS.
OC       BACTERIA: FIRMICUTES: ACTINOBACTERIA: ACTINOBACTERIALE:
OC       ACTINOMYCETALES: CORINNEBACTERIINEAE: MYCOBACTERIACEAE: MYCOBACTERIUM.
RN       [1]
RP       SEQUENCE FROM N.A.
RC       STRAIN-H37RV;
RA       BADOCK K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., PARKHILL J.;
RL       SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBT DATA BANKS.
CC       - FUNCTION: CELL WALL FORMATION (BY SIMILARITY).
CC       - I- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL + L-ALANINE -
CC       ADP + ORNITHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.
CC       - PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC       - SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC       - SIMILARITY: BELONGS TO THE MURCDFE FAMILY.
-----
Cc       This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc       between the Swiss Institute of Bioinformatics and the EMBL Outstation -
Cc       the European Bioinformatics Institute. There are no restrictions on its
Cc       use by non-profit institutions as long as its content is in no way
Cc       modified and this statement is not removed. Usage by and for commercial
Cc       entities requires a license agreement (see http://www.isb-sib.ch/announce/
Cc       or send an email to license@isb-sib.ch).
-----
DR       EMBL: Z95388; E315954; -
KM       PEPTIDOGLYCAN SYNTHESIS: CELL WALL; CELL DIVISION; LIGASE;
KM       ATP-BINDING.
FT       NP_BIND     122      128      ATP (POTENTIAL).
SQ       SEQUENCE 494 AA: 51176 MW: 841D1343 CRC32:

Query Match           11.2%: Score 349; DB 1; Length 494;
Best Local Similarity 28.9%; Pred. No. 8,50e+40;
Matches 136; Conservative 107; Mismatches 185; Indels 42; Gaps 31;

Db      14 HMVGIGAGMSGIARILLDRGGLVSGSDAKESRGVH-ALRARGALIRIGHDASSDLLPG 72
QY      5 HGVGIGGSMSSLAQIMDMDLGHEVOGSDLENVTFLEVALLRNKG--IKLIPGAN-NI-KE 60
Db      73 GATAVVTAHPKTNPELVARRRGIPVULRAPVALIKMGRTTLMTVTHGKTITTSM 132
QY      61 D-MVVYIOGNA-FASHEEIVEYRAHQKLTVVSNDQLGDIIDYTSVAVTAGHGKSTTGL 118
Db      133 LIVAAOHGGLDPSFVAGSELGEAGTNAHHGSDCVVAADAEDSGSLLOYTPHAVYITNE 192
QY      119 LSHVN-NDDKRTSLFIIG-D-G-TGGLPE-S-DYFAFACEYRRHFLSYKPDYALMTND 172

```







Mon Nov 29 10:16:56 1999

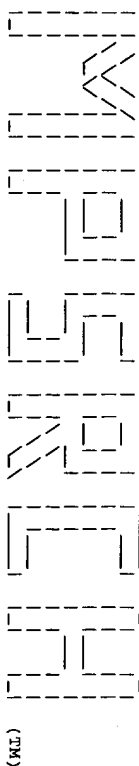
US-09-103-287-2.rsp

Page 10

Db 307 TNFSGVPHR 315  
|:|:|  
Oy 290 ETEGSKRR 298

Search completed: Wed Nov 24 00:32:46 1999  
Job time : 92 secs.





Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Nov 24 00:33:03 1999; Maspar time 27.28 Seconds

Tabular output not generated. 874.409 Million cell updates/sec

Title: >US-09-103-287-2  
Description: (1-437) from US09103287.pep  
Perfect Score: 3121  
Sequence: 1 MTHYFVGIKSGMSLSAQI.....GDIOKLQNAVLDKLGKNAF 437

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database:

sptrembl9  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle  
9:sp-phage 10:sp-plant 11:sp-prodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 49.820; Variance 102.447; scale 0.486

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	620	19.9	454	2	UDP-N-ACETYLMURAMATE-A	1.62e-91
2	567	18.2	803	2	UDP-N-ACETYLMURAMATE-A	1.19e-81
3	202	6.5	235	2	UDP-N-ACETYLMURAMATE-A	1.81e-14
4	154	4.9	682	1	UDP-N-ACETYLMURAMATE-A	5.01e-07
5	128	4.1	445	2	UDP-MURNA-C-PENTAPEPTIDE	2.31e-03
6	124	4.0	856	2	HYPOHETICAL 99.3 KD P	7.88e-02
7	119	3.8	482	2	UDP-MURNA-C-TRIPETIDE	2.63e-02
8	120	3.7	354	5	598AA LONG HYPOHETICAL	8.53e-02
9	116	3.7	354	5	42 KDA TRANSPOSASE.	1.52e-01
10	114	3.7	354	5	42 KDA PROTEIN.	6.37e-02
11	117	3.7	492	2	UDP-N-ACETYLMURAMYL-TR	1.52e-01
12	114	3.7	571	2	DEOXYRIBODIPYRIMIDINE	3.59e-01
13	111	3.6	121	13	THYROID HORMONE RECEPT	3.59e-01
14	111	3.6	272	2	TRANSCRIPTIONAL ACTIVA	3.59e-01
15	111	3.6	350	2	UDP-N-ACETYLMURAMYL-TR	3.59e-01
16	111	3.6	490	2	UDP-N-ACETYLMURAMYL-TR	3.59e-01
17	111	3.6	493	2	UDP-MURNA-C-PENTAPEPTIDE	3.59e-01
18	112	3.6	517	2	UDP-N-ACETYLMURAMYL-TR	2.70e-01
19	108	3.5	143	1	MOLYBDENUM COFACTOR BI	8.33e-01
20	108	3.5	317	10	AP2 DOMAIN CONTAINING	8.33e-01

21	109	3.5	585	2	066671	HYPOHETICAL 68.4 KD P	6.31e-01
22	108	3.5	1161	2	069276	NONTOXIC-NONHAEMAGGLUT	8.33e-01
23	108	3.5	1161	2	045891	NNNH PROTEIN.	8.33e-01
24	108	3.5	1193	2	045880	NNNH.	8.33e-01
25	108	3.5	1198	2	069277	NONTOXIC-NONHAEMAGGLUT	8.33e-01
26	108	3.5	1198	2	045893	NNNH PROTEIN.	8.33e-01
27	110	3.5	1161	5	045497	SUR-2 PROTEIN.	4.76e-01
28	110	3.5	1671	1	P77933	DNA-DEPENDENT DNA POLY	4.76e-01
29	105	3.4	245	1	028187	HYPOHETICAL 28.6 KD P	1.90e+00
30	107	3.4	321	2	056873	WBCJ PROTEIN (ORF1.8)	1.10e+00
31	105	3.4	325	5	002304	TO9E11.1 PROTEIN.	1.10e+00
32	106	3.4	354	5	P90561	42 KDA 'D,D35E' TRANSP	1.45e+00
33	107	3.4	371	2	050503	PERIPLASMIC SOLUTE-BIN	1.10e+00
34	106	3.4	374	2	P74121	HYPOHETICAL 42.9 KD P	1.45e+00
35	106	3.4	447	2	026027	UDP-MURNA-C-TRIPETIDE	1.45e+00
36	106	3.4	2457	5	077375	M1366.20 PROTEIN	1.45e+00
37	103	3.3	429	2	051747	CONSERVED HYPOHETICAL	3.27e+00
38	102	3.3	436	2	052187	IGG-BINDING PROTEIN SB	4.27e+00
39	102	3.3	468	2	059822	DIHYDROLIPOMIDE DEHYD	4.27e+00
40	104	3.3	540	1	033398	MINOR STRUCTURAL GAS V	2.49e+00
41	104	3.3	1014	2	047107	CYTOTOXIC NECROTIZING	2.49e+00
42	103	3.3	1928	4	099736	HGGC1 (FRAGMENT)	3.27e+00
43	103	3.3	2412	4	092616	MELDEBLAST KIA0219 (F	3.27e+00
44	104	3.3	2919	14	085431	RNA POLYMERASE.	2.49e+00
45	102	3.3	3078	5	026031	VARIANT-SPECIFIC SURFA	4.27e+00

#### ALIGNMENTS

RESULT ID	1	PRELIMINARY;	PRT;	454 AA.
AC	067373:			
DT	01-AUG-1998 (TREMBLREL. 07, CREATED)			
DT	01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	UDP-N-ACETYLMURAMATE-ALANINE LIGASE.			
GN	MURC.			
OS	AQUIFEX AEOLICUS.			
OC	BACTERIA; AQUITICILES; AQUITICACEAE; AQUIFEX.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRATIN-VF5;			
RC	MEDLINE: 98196666.			
RA	DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,			
RA	GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AJAYI M., HUBER R.,			
RA	FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;			
RT	"The complete genome of the hyperthermophilic bacterium Aquifex			
RT	aeolicus.";			
RL	NATURE 392:353-358(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRATIN-VF5;			
RA	DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,			
RA	GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AJAYI M., HUBER R.,			
RA	FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;			
DR	EMBL: A0000736; G2983764; .			
KW	LIGASE.			
SQ	SEQUENCE 454 AA; 50893 MW; 120799CE CRC32;			
Query Match	19.9%; Score 620; DB 2: length 454;			
Best Local Similarity	30.8%; Pred. No. 1.62e-91;			
Matches	139; Conservative 125; Mismatches 164; Indels 24; Gaps 20;			
Db	6 IKKFFHIGIGIGMSGIAQLLEMGYKVSQSDISENKKTKL-LKQGAQYIIGHREPNIG 64			
Qy	1 MTHYFVGIKSGMSLSAQIMHDLGHEVQSGSDIENYVFTEVALRNNGIKI-LFGANNIK 59			
Db	65 DAQVYVSSAVKNDNPDIQAKRRNIPVPRGEMALFLFKAGIAGVSGHGKTTTSMI 124			
Qy	60 EDWVVIQGNFASHSHEIYAHQIKLDVVSYNDFLQIIDQYTSVAVTGAHGKSTGTGL 119			
Db	125 AEILINAGLEPYIIGIRLRLKLTGNKLGELGELVSEADESDSFLKLPRAVAIVINVK 184			



```

QY      120 SHWM-VGDKRTSLID-----GTGKLPESDFAFACERYRRHFSYKPDYAIMNIDF 173
      185 EHLDENEFERNVEAFEOFNNSVPYFGFAVYNLDDEPTLQLVKSKSHRYVTGINSPLY 244
      174 DHDYIKKIDNDVDAIOEAMHNKK-GI-IAWGDEHLK-IE-ADVPITYYFKKSDDI 229
      245 RAKNLYKGRVEFEVFKGKELRHLGTA-GIHVYNALATGVA-LE-LGVSEVIR 301
      230 YKONIGITDKGFADYVYDGEFDFH-LSPQYGDHTVYNALAVIAISYLEKLDVT-N-1K 286
      302 KSLSEERNARRELKGYKNSPYDYDGHHPFEIKAVINSLDMPDKNLVFPQPHRY 361
      287 EALETGGVKRRRENETTI-ANQYIVDDYAHHPREISATIDTARKKYPHKEVAVFPQPHF 345
      362 SRTYLFEDFVKYKIDKLVTDIYPASENNYVGSAEELARK-SGAVFANKDEVEPK 420
      346 SRTQAFLEAFESLCKADRYFLCEITGSIRESNGALFTIIDLKIGASFI-N-EDLIN- 402
      421 VREHDEGVILFLGAGSISKWCEELKEVNL 452
      403 VLEQFDNA-VVLEPMGAGDIQKQNAVILDKLGM 433

```

```

RESULT 2
ID 084767 PRELIMINARY: PRT: 803 AA.
AC 084767;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
GN UDP-N-ACETYLMURAMATE-ALANINE LIGASE AND D-ALA-D-ALA LIGASE.
OC MURC/DDLA.
OS CHLAMYDIA TRACHOMATIS.
OC BACTERIA; CHLAMYDIALES; CHLAMYDIACEAE; CHLAMYDIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UM-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATISOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT Chlamydia trachomatis.";
RL Science 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UM-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATISOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RT SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AE001348; G3329224; -.
KW LIGASE.
SQ SEQUENCE 803 AA; 89230 MW; 7DB58969 CRC32;

```

Query Match 18.2%; Score 567; DB 2; Length 803;  
 Best Local Similarity 29.9%; Pred. No. 3.19e-81;  
 Matches 120; Conservative 106; Mismatches 162; Indels 14; Gaps 13;

```

Db 7 YHFGIGGIGMSALAHVLLDRGYSVSGSDLSSEKRYE-RLKNGAFELFGNDEHPEEA 65
      4 YHFGVIGKSGMSLAQIMHDLGHEVGSIDENVTFEVALRNKGIRI-LPFGANNKEDM 62
      66 VVVYSSISKKNPEFLSAKSGNRVVRRAELLAELADQISIFVTSKGTYSILTAI 125
      63 VVIOGNAFASHEIYRAHQKLDVSYNDPLGOIIDQYTSVAVTGAHKGITSTGLSHV 122
      126 LQAKKNPFAIGLQNEINGSGSEYFVAEADSDGSIKCTYTPESVITNIDHELSN 185
      123 MNGDKKT-SFLIGD-GT-GM-GLPESDYFAFEACEYRRHFLSYKPDYAIMTNIDFDPDY 178
      186 FEGDRELLASLKDFA--LKTQOICWYNGDCPR-LNSCLQGHFGDSCDLHLSYGE 242
      179 FQDINDVF-DAFOEMAHNVKGIITIAWGDEHLKIEADVPITYYFGKSDSDIYKONIQIT 237

```

```

Db 243 GWRLEYTAYQVQV-YADIEVQVGMHVNLAAMGIALSLGIDEGAIIRNAFRGSSVQ 301
      238 DGTAFDV-YVNGEYFDHFLSPQYGDHTVYNALAVIAISYLEKLDVTINKELTFGGVK 296
      302 RRLQKNSSEIFLEIDYAHHPSEISCTLRRAVRAVAGORRIATIQPHRSRLRECIOSF 361
      297 RRFNETTIANQ-VIYDDVAHHPREISATIDTARKKYPHKEVAVFPQPHFSRTQAFLENF 355
      362 PSAPFADDEVILTEVYSAG-EEADISYOKLAELAISQSIYK 402
      356 ASLCKADRVFLCEITGSIRESNGALFTIIDLKIGASFIN 397

```

```

RESULT 3
ID 069551 PRELIMINARY: PRT: 235 AA.
AC 069551;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE UDP-N-ACETYLMURAMATE-ALANINE LIGASE (FRAGMENT).
GN MURC.
OS MYCOBACTERIUM LEPRAE.
OC BACTERIA; FIRMITICES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; CORYNEBACTERIINAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RA BROWN D., CHURCHER C.M.;
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93186700.
RA EIGMEIER K., HONORE N., WOODS S.A., CAUDRON B., COLE S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization
RT of Mycobacterium leprae.";
RL MOL. MICROBIOL. 7:197-206(1993).
DR EMBL; AL022602; E1287791; -.
KW LIGASE.
FT NON_TER 1
FT NON_TER 235
SQ SEQUENCE 235 AA; 24350 MW; 708A69C2 CRC32;

```

Query Match 6.5%; Score 202; DB 2; Length 235;  
 Best Local Similarity 32.7%; Pred. No. 1.81e-14;  
 Matches 66; Conservative 47; Mismatches 73; Indels 16; Gaps 12;

```

Db 14 HMVIGGSGMSIARILLDRGLVSGSDAKESRYVH-ALRAGALIRIGHDASLIDLPG 72
      5 HFGVIGKSGMSLAQIMHDLGHEVGSIDENVTFEVALRNKGIRI-LPFGAN--NT-KE 60
      73 GATAVITRTAIPTKPNPELVEARRRGIPVLLRSAYLARIMDGCTLWVAGTGHGKTTTSM 132
      61 D-MVVIQG-NAFASHEIYRAHQKLDVSYNDPLGOIIDQYTSVAVTGAHKGITSTGL 118
      133 LVVALQHGCGPSFVYGGELAVVGTNAHNGJACGVAADESDGSLLETYNVAVVTNID 192
      119 LSHVN-NQDKKTSFLIG-D-GT-GMGLPES--DYFAFEACEYRRHFLSYKPDYAIMTNID 172
      193 SDHLDFGVSVDAYIRVPDSFE 214
      173 FHPDYFKDIND--VFDAFOE 191

```

```

RESULT 4
ID 026630 PRELIMINARY: PRT: 682 AA.
AC 026630;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

```



DE UDP-N-ACETYLURAMYL TRIPEPTIDE SYNTHETASE RELATED PROTEIN.  
GN MTH30.  
OS METHANOBACTERIUM THERMAUTOTROPICUM.  
OC ARCHAEA: EURYARCHAEOTA; METHANOBACTERIALES; METHANOBACTERIACEAE;  
OC METHANOBACTERIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DELTA H;  
RX MEDLINE: 98037514.  
RA SMITH D.R., DOUGHERTY STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,  
RA ALDENDE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,  
RA HARRISON D., HOANG L., KEAGLE P., LUM W., POTTER B., QIU D.,  
RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,  
RA JIMANI N., CARUSO A., BUSH D., SAHER H., PATWEL D., PRABHAKAR S.,  
RA MCDUGALL S., SHIMER G., GOYAL A., PITROVSKI S., CHURCH G.M.,  
RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REVE J.N.:  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
RT delah: functional analysis and comparative genomics."  
RL J. BACTERIOL. 179:7135-7155(1997).  
DR EMBL: AE000836; G2621604;  
SQ SEQUENCE 682 AA; 73795 MW; 67377BPE CRC32;  
Query Match 4.98; Score 154; DB 1; Length 682;  
Best Local Similarity 30.98; Pred. No. 5.01e-07;  
Matches 30; Conservative 21; Mismatches 44; Indels 2; Gaps 2;  
DB 484 MAIEASPGGFKVIGGEMREVLATP-GIFVNYNALAATATMGLEIDDIYRGLESFK 542  
QY 234 IQITDKGTFADYVGDGFDFHLSPOYGDHTVNALAVIAISLEKLDVTINKEALFTG 293  
DB 543 GVPGRFQELSEPRILLD-YAHNPAGVRAVMODLRK 578  
QY 294 GVKRRNETTIANQVIVDDYAHHPREISATIDTARK 330  
RESULT 5  
ID 067002 PRELIMINARY; PRT; 445 AA.  
AC 067002;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE UDP-MURNAC-PENTAPEPTIDE SYNTHETASE.  
GN MURE.  
OS AQUIFEX AEOLICUS.  
OC BACTERIA: AQUIFICALES; AQUIFICACEAE; AQUIFEX.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-VF5;  
RX MEDLINE: 98196666.  
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.:  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
RT aeolicus."  
RT NATURE 392:353-358(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-VF5;  
RX DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.:  
RT SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AE000709; G2983375;  
SQ SEQUENCE 445 AA; 49581 MW; 51EEC831 CRC32;  
Query Match 4.18; Score 128; DB 2; Length 445;  
Best Local Similarity 26.28; Pred. No. 2.31e-03;  
Matches 28; Conservative 33; Mismatches 38; Indels 8; Gaps 6;  
DB 235 SFGRELEAONAVESEKVSFEV--EGKFTISVPSLIVEN-ILACCYLVK--LG-F 288  
QY 222 GKRSDDIYANQIQTGTATFDVYDGE-FYDHLSPQYGDHTVNALAVIAISYLEK 280

DB 289 TLKEFKALKSESPVEGRMKVFNKFNILLIDPTVYANPSLRNAIKS 335  
QY 281 DVTNIKEALFTGKVRKRENETTIANQVIVDD-YAHHPREISATIDT 326  
RESULT 6  
ID 025974 PRELIMINARY; PRT; 856 AA.  
AC 025974;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 99.3 KD PROTEIN.  
GN HELI433.  
OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).  
OC BACTERIA: PROTEOBACTERIA: EPSILON SUBDIVISION; HELICOBACTER GROUP;  
OC HELICOBACTER.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-26695;  
RX MEDLINE: 97394467.  
RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,  
RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,  
RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,  
RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLOBE A.,  
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,  
RA BERG D.E., GOCCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,  
RA COTTON M.D., WEIMAN J.M., FUJII C., BOWMAN C., MATTHEY L., WALLIN E.,  
RA HAYES W.S., BORODOVSKY M., KAR P.D., SMITH H.O., FRASER C.M.,  
RA VENTER J.C.:  
RT "The complete genome sequence of the gastric pathogen Helicobacter  
RT pylori[unpublished erratum appears in Nature 1997 Sep  
RT 25:389(6649):412]."  
RL NATURE 388:539-547(1997).  
DR EMBL: AE000643; G2314608;  
DR TIGR: HP1433;  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 856 AA; 99278 MW; 3D9A8B88 CRC32;  
Query Match 4.08; Score 124; DB 2; Length 856;  
Best Local Similarity 26.48; Pred. No. 7.88e-03;  
Matches 38; Conservative 35; Mismatches 61; Indels 10; Gaps 10;  
DB 108 VPD-FKTSQCEKEFKFNKLNSSHLKSIQIDKNNFFIYQWLEIVKPTIDIMNEVANT 166  
QY 185 VPDAGQEMAHNKKGIANGDDEHL-RKTEADVPYIYFKSDDIYANQIQT-DKGT 242  
DB 167 KDI-LDADYYLADLIS-D-GDKTIEKLTIRSSHYKLN-RGVNE-LCKMDFMEYGF 221  
QY 243 FVYVYDGEFY-DHFLSPQYGDHTVNALAVIAISYLEKLDVTNIKEALFTFGVKRRFE 301  
DB 222 SQQAHQEFMSYVERPPKREFQASI 245  
QY 302 TITANQVIVDDYAHHP-REISATI 324  
RESULT 7  
ID 067631 PRELIMINARY; PRT; 482 AA.  
AC 067631;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE UDP-MURNAC-TRIPETIDE SYNTHETASE.  
GN MURE.  
OS AQUIFEX AEOLICUS.  
OC BACTERIA: AQUIFICALES; AQUIFICACEAE; AQUIFEX.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-VF5;  
RX MEDLINE: 98196666.  
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.:  
RT "The complete genome of the hyperthermophilic bacterium Aquifex



```

RT aeolicus.
RL NATURE 392:353-358(1998).
RN 121
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RA DECKERT G., WARREN P.V., GASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AULAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AE000754; G2984050;
SO SEQUENCE 482 AA; 53991 MW; A3154DEB CRC32;

Query Match
Best Local Similarity 27.2%; Score 119; DB 2; Length 482;
Pred. No. 3.53e-02;
Matches 52; Conservative 40; Mismatches 85; Indels 14; Gaps 14;

Db 183 QYRYPTFEAVLFTNLSODHLDYKHTMEDYFASAKAKLFTYESVKIINADDTYGRKLL 242
QY 157 HFLSKPDY-ILM-TNIDFDPDKINDVDFAQEM-AHNVKGIILANGDEHLRRI- 212
Db 243 KITHEILTYG-KKGD-LKILNFRTPFGSALRIAFKGEYE-FSTNLGDFQAVN-LSA 298
QY 213 E-ADVPITYYGFSDSDIYAQNIQITDGTAFDVYVDEDFYHFLSPQYGDHTVINALAV 271
Db 299 -AIAVALMKGIIEPDILQALKCVN-VPGREYVSHSKFTVIIDYAHTPDAVENLRTARK 356
QY 272 IAI5Y-LKK-LDVTNIKALETFEGVKRRFNETTIANOVIVDYVAHHRFISATIDTARK 329
Db 357 -LSKGLISVF 366
QY 330 KYPKREVAVP 340

RESULT 8
ID 058475. PRELIMINARY: PRT: 598 AA.
AC 058475.
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE 598AA LONG HYPOTHETICAL PROTEIN.
GN PH0785.
OS PYROCOCUS HORIKOSHII.
OC ARCHAEA: EURARCHAEOTA: THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCUS.
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
RX MEDLINE: 98344137.
RA KANARABAYAST Y., SAMADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSORYAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMITA M., OHFUKU Y.,
RA FUJIMASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIUYA H.,
RA KIRUCHI H.;
RT "Complete Sequence and Gene Organization of the Genome of a
RT Hyper-Thermophilic Archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA RES. 5:55-76(1998).
DR EMBL: AP000003; D1030821;
SO SEQUENCE 598 AA; 68436 MW; 47391D5E CRC32;

Query Match
Best Local Similarity 32.8%; Score 120; DB 1; Length 598;
Pred. No. 2.63e-02;
Matches 19; Conservative 17; Mismatches 19; Indels 3; Gaps 3;

Db 516 FSTSEFRIERISGKGRALIEYLDGEFTIKRYEFKREYGERIVLVKVRDPGMYL 573
QY 223 FKSDSDIYAQNIQITDGTAFDVYVDEDFYHFLSPQYGD-HTVINALAVIALSYL 277

RESULT 9
ID P90564. PRELIMINARY: PRT: 354 AA.
AC P90564;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)

```

```

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE 42 KDA TRANSPOSASE.
OC OXYTRICHA FALLAX.
OS EUKARYOTA: ALVEOLATA: CILIOPHORA: HYPOTRICHs: STICHOTRICHIDA;
OC OXYTRICHIDAE: OXYTRICHA.
RN 111
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TBE1 FAL4;
RX MEDLINE: 86079559.
RA HERRICK G., CARTINOUR S., DAMSON D., ANG D., SHEETS R., LEE A.,
RA WILLIAMS K.;
RT "Mobile elements bounded by C4A4 telomeric repeats in Oxytricha
RT fallax.";
RL CELL 43:759-768(1985).
RN 121
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TBE1 FAL4;
RX MEDLINE: 94038943.
RA WILLIAMS K., DOAK T.G., HERRICK G.;
RT "Developmental precise excision of Oxytricha trifallax
RT telomere-bearing elements and formation of circles closed by a copy
RT of the flanking target duplication.";
RL EMO J. 12:4593-4601(1993).
RN 131
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TBE1 FAL4;
RX MEDLINE: 94134747.
RA DOAK T.G., DOERDER F.P., JAHN C.L., HERRICK G.;
RT "A proposed superfamily of transposase genes: transposon-like
RT elements in ciliated protozoa and a common "D35E" motif.";
RL PROC. NATL. ACAD. SCI. U.S.A. 91:942-946(1994).
RN 141
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TBE1 FAL4;
RA DOAK T.G., WITHERSPOON D.J., DOERDER F.P., DR WILLIAMS K.,
RA HERRICK G.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U85403; G1815756;
SO SEQUENCE 354 AA; 41984 MW; DEBA736B CRC32;

Query Match
Best Local Similarity 24.1%; Score 116; DB 5; Length 354;
Pred. No. 8.53e-02;
Matches 20; Conservative 27; Mismatches 33; Indels 3; Gaps 3;

Db 127 YKRNIGTYKGVNRKRYIDFTQPFYRNANLYYDQMORAQONNIK-FIH-ADEAVF-TFN 183
QY 154 YRRFLSKPDYALMTNIDFDPDKINDVDFAQEMAHNVKGIILANGDEHLKRIE 213
Db 184 TFIKSMYKRNSTNIEYVDQYKV 206
QY 214 ADVPIYTYGFDSDIYAQNIQI 236

RESULT 10
ID Q94991. PRELIMINARY: PRT: 354 AA.
AC Q94991;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE 42 KDA PROTEIN.
OC OXYTRICHA FALLAX.
OS EUKARYOTA: ALVEOLATA: CILIOPHORA: HYPOTRICHs: STICHOTRICHIDA;
OC OXYTRICHIDAE: OXYTRICHA.
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-SOBKARYONIDE 3.5;
RX MEDLINE: 88118409.
RA HERRICK G., CARTINOUR S.W., WILLIAMS K.R., KOTTER K.P.;
RT "Multiple sequence versions of the Oxytricha fallax 81-MAC alternate
RT processing family.";
RL J. PROTOZOOL. 34:429-434(1987).
RN 121
RP SEQUENCE FROM N.A.

```



```

RL MICROR. NRUG RES. 4:85-90(1998),
SR EMBL: Y14370. E1301684. -
SQ SEQUENCE 492 AA; 54099 MW; 7E0C9C3B CRC32;

Query Match 3.7%; Score 117; DB 2; Length 492;
Best Local Similarity 27.6%; Pred. No. 6,37e-02;
Matches 35; Conservative 34; Mismatches 47; Indels 11; Gaps 10;

Db 245 EYLRTV-TPEVEYSYGDDEAQAQMAKIQESLQGVSPD-FVTP-FGTYPYKSPYVKGKFN 301
OY 207 EHLKRITADVPYIYGGFKSDSDDIYAQNIDITDKGTADYVDGGEFYDHPFLSPQY-CDHTV 265
Db 302 SNMAAMAVNSKSTSL-ETIK-AVENLEPVEBERLEVLDPSPLEIDLTID-VYATAGMN 358
OY 266 LNALA-YVAISYLEK-LDVTNINKEALETFEGVAKRRN--ETTIANOVYVDYDAHHPREIS 321
Db 359 KLIDAVQ 365
OY 322 ATIDTAR 328

RESULT 12
ID 084898 PRELIMINARY; PRT; 571 AA.
AC 084898;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE DEOXYRIBODIPYRIMIDINE PHOTOLYASE.
GN UVRC.
OS MYCOPLASMA BOVIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PG45;
RX MEDLINE: 98339280.
RA SUBRAMANIAM S., BENGONIER D., POUMARAT F., CAPAUL S., SCHLATTER Y.,
RA NICOLET J., FREY J.;
RT "Species identification of Mycoplasma bovis and Mycoplasma agalactiae
RT based on the uvrC genes by PCR.";
RL MOL. CELL. PROBES 12:161-169(1998).
DR EMBL, AF003959; G3435082; -.
KW LIASE.
SQ SEQUENCE 571 AA; 66463 MW; 75E16439 CRC32;

Query Match 3.7%; Score 114; DB 2; Length 571;
Best Local Similarity 27.2%; Pred. No. 1.52e-01;
Matches 22; Conservative 25; Mismatches 27; Indels 7; Gaps 7;

Db 271 VDESIRVFEQFYE-DKILPDNLIVQELLNFDNLSESE-YK-FISPKIGTINKKVD-LA 326
OY 212 IEADVPIIYCYFKSDSDDIYAQNIDITDKGTAFDVYDGEFYDHPFLSPQY-DHTVLNALA 270
Db 327 ILINDIYEKEHLV-INOLD 346
OY 271 VLAIS-YLEKLDVTNINKEALE 290

RESULT 13
ID 091003 PRELIMINARY; PRT; 121 AA.
AC 091003;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE THYROID HORMONE RECEPTOR BETA 2 (FRAGMENT).
GN CTR BETA 2.
OS GALLUS GALLUS (CHICKEN).
OC EUAROTOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAUROTA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-EYE;
RX SJOEBERG M., VENNSTROEM B., FORREST D.;

```



RL DEVELOPMENT 114:39-47(1992).  
DR EMBL: X62642; G63823; -.  
KW THYROID HORMONE.  
FT NON\_TER 121 121  
SQ SEQUENCE 121 AA; 14088 MW; B08AA3D9 CRC32;  
Query Match 3.6%; Score 111; DB 13; Length 121;  
Best Local Similarity 25.8%; Pred. No. 3.59e-01;  
Matches 16; Conservative 19; Mismatches 22; Indels 5; Gaps 5;  
Db 62 TCSPGN-V-FKSESDMDMLNOYNOPEYFTEEPFTEESQVSPSYOKKGYPSYLDKD 119  
Y 151 ACEYRRHFLSYKPDYAMT-NID-FDHPDYFKINDVDFADQEMAHNVKGIIT-AMGDDE 207  
Db 120 EL 121  
Y 208 HL 209  
RESULT 14  
ID 007681 PRELIMINARY; PRT; 272 AA.  
AC 007681;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)  
DE TRANSCRIPTIONAL ACTIVATOR AFRR.  
GN AFRR.  
OS ESCHERICHIA COLI.  
OC PLASMID UNDESIGNATED.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
OC ESCHERICHIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RDEC-1;  
RX MEDLINE; 90202141.  
RA WOLF M.K., BOEDEKER E.C.;  
RT "Cloning of the genes for AF/R1 pilin from rabbit enteroadherent  
RT Escherichia coli RDEC-1 and DNA sequence of the major structural  
RT subunit".  
RL INFECT. IMMUN. 58:1124-1128(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RDEC-1;  
RA CANTLEY J.R., MOSELEY S.L.;  
RT "Characterization of the Escherichia coli AF/R1 pilus operon: novel  
RT genes necessary for transcriptional regulation and pilus mediated  
RT adherence".  
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AF050217; G3372510; -.  
DR PFAM: PF00165; HTH\_2; 1.  
KM PLASMID.  
FT CONFLICT 58 58 R -> S (IN REF. 1).  
SQ SEQUENCE 272 AA; 32359 MW; 559E9E09 CRC32;  
Query Match 3.6%; Score 111; DB 2; Length 272;  
Best Local Similarity 25.0%; Pred. No. 3.59e-01;  
Matches 24; Conservative 27; Mismatches 37; Indels 8; Gaps 7;  
Db 1 MKLTLTINATQYIDDNESISIDINALVDYSGYSRRLQLLFKENIGVTIGKTYQLRRI 60  
Y 265 VLNALAVIAIS-YLE-KLDVNTIK-EALETFGKVRRENETTI-AN-OVIYDYAHHPRE 319  
Db 61 TBAAT-LLR--FTNLKIVDISERLFYDSQOTFREF 93  
Y 320 ISATIDTARKKYPHKEVAVFQPHTFSTQALNEF 355  
RESULT 15  
ID 069290 PRELIMINARY; PRT; 350 AA.  
AC 069290;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMATE-- 2,6-DIAMINOPIMELATE LIGASE  
(FRAGMENT).  
DE (FRAGMENT).  
GN MURE.  
OS CAMPYLOBACTER JEJUNI.  
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; CAMPYLOBACTER GROUP;  
OC CAMPYLOBACTER.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NCTC 11168;  
RA GRIFFITHS P.L., CONNERTON I.F.;  
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: Y16882; E1266307; -.  
KW LIGASE.  
FT NON\_TER 350 350  
SQ SEQUENCE 350 AA; 38656 MW; 0E4D2922 CRC32;  
Query Match 3.6%; Score 111; DB 2; Length 350;  
Best Local Similarity 29.2%; Pred. No. 3.59e-01;  
Matches 33; Conservative 26; Mismatches 48; Indels 6; Gaps 5;  
Db 227 FTYGI-ENPALYQIKAYSLEEGISTIVTKNKQTF-HIDSPILGLFNLVNLVASCYNEL 284  
Y 219 YTFKDSDDIYAQNIOTDKTAFDYYVDGEYDHLSPYGDHVLNALAVIA-TSYL 277  
Db 285 VKPDLKDLKKAISGFGVCGRVEQ--YAKGIYD-FAHTPDGIEKVLDTLKNK 334  
Y 278 EKLVTNKEALETFEGVKKRRFNETTIANQYIVDYVAHHPREISATIDTARKK 330  
Search completed: Wed Nov 24 00:36:17 1999  
Job time : 194 secs.







Seq primer: M13 Forward.  
Location/Qualifiers

FEATURES  
source

1..385

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/note="Vector: p773d-Pac (Pharmacia) with a modified  
polylinker. Site 1: Not I; Site 2: Eco RI; The UI-R-E1  
library is a subtracted library derived from the UI-R-E0  
library. The UI-R-E0 library consisted of a mixture of  
individually tagged normalized libraries constructed from  
8, 12 and 18-day embryo. The tag is a string of 3-5  
nucleotides present between the Not I site and the  
oligo-dT track which allows identification of the library  
of origin of a clone within the mixture. The subtracted  
library (UI-R-E1) was constructed as follows: PCR  
amplified cDNA inserts from a pool of UI-R-E0 clones from  
which 3 ESTs had been derived was used as a driver in a  
hybridization with the UI-R-E0 library in the form of  
single-stranded circles. The remaining single-stranded  
circles (subtracted library) was purified by  
hydroxyapatite column chromatography, converted to  
double-stranded circles and electroporated into DH10B  
bacteria (Life Technologies) to generate the UI-R-E1  
library. This procedure has been previously described  
(Bonaldi, Lennon and Soares, Genome Research 6: 791-806,  
1996)"

/db\_xref="taxon:10116"

/map="MMU16C3-C4 region: MMU16C3-C4 region"

/clone="UI-R-E1-f-e-02-0-UI"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies)"

/base\_count 72 a 113 c 101 g 99 t

Query Match 3.0%; Score 162; DB 23; Length 385;

Best Local Similarity 34.7%; Pred. No. 1.25e-07;

Matches 42; Conservative 29; Mismatches 50; Indels 0; Gaps 0;

249 AAGGAAGCTCGCTGATTCAGGCGGCGCCCTGCTGACACCTTGG 308

174 HAARGARGAATGATGATGATGATGATGATGATGATGATGATGAT 233

309 GCTCATGCTCCAGCTCACTGATGATGATGATGATGATGATGATGAT 368

234 NMGNCNCATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 293

369 C 369

294 H 294

RESULT 2

LOCUS AF029557 903 bp DNA GSS 09-JAN-1998

DEFINITION Saimonella typhimurium strain LT2 clone 991-T3, LI-Cor, genomic

ACCESSION AF029557

VERSION 92571087

KEYWORDS AF029557.1 GI:2571087

SOURCE GSS.

ORGANISM Saimonella typhimurium.

Eubacteria: Proteobacteria: gamma subdivision; Enterobacteriaceae;

Saimonella.

1 (bases 1 to 903)

Wong, R.M.Y., Benson, N. and McClelland, M.

End sequences of Saimonella typhimurium strain LT2 lambda DASH II

clones, LI-Cor

Unpublished

2 (bases 1 to 903)

Wong, R.M.Y., Benson, N. and McClelland, M.

Direct Submission

Submitted (10-OCT-1997) Molecular Biology, Sidney Kimmel Cancer

Center, 3099 Science Park Road, San Diego, CA 92121, USA

FEATURES  
source

1..903

/organism="Saimonella typhimurium"

/strain="LT2"

/note="end sequence from lambda clone"

/db\_xref="taxon:602"

/clone="991-T3, LI-Cor"

BASE COUNT 178 a 219 c 265 g 217 t 24 others

Query Match 3.0%; Score 165; DB 37; Length 903;

Best Local Similarity 32.6%; Pred. No. 2.77e-08;

Matches 115; Conservative 53; Mismatches 185; Indels 0; Gaps 0;

428 CACATTCACCTTTGCGGATCGGTCGTCGTCGTCGTCGTCGTCGTCG 487

7 CATTTCATTTTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 66

488 AATGAAGGCTATTCAGATCGATTCGATTCGATTCGATTCGATTCGAT 547

67 GATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 126

548 ACGAGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 607

127 YTNMGNAAYAAAGGATTAATTAATTAATTAATTAATTAATTAATTA 186

608 GTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 667

187 GTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 246

668 GCGGCTATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 727

247 YTNMGNAAYAAAGGATTAATTAATTAATTAATTAATTAATTAATTA 306

728 GGCATCGCATTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780

307 WSNGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 359

RESULT 3

LOCUS AF034173 2275 bp mRNA EST 30-MAR-1998

DEFINITION AF034173 Human mRNA (Tripodis and Ragousis) Homo sapiens CDNA

ACCESSION AF034173

VERSION 92707735

KEYWORDS AF034173.1 GI:2707735

SOURCE EST.

ORGANISM human.

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia;

Eutheria: Primates: Catarrhini: Hominiidae: Homo.

1 (bases 1 to 2275)

Tripodis, N. and Ragousis, J.

Generation of a transcription map in the region immediately

centromeric to human MHC across the 6p21.2-6p21.3 chromosomal

boundary.

Unpublished (1997)

On Jan 19, 1998 this sequence version replaced gi:2045115.

Center, 3099 Science Park Road, San Diego, CA 92121, USA

FEATURES  
source

1..2275

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="6p21.3"

/clone="ntcon2 contig"

BASE COUNT 438 a 619 c 470 g 599 t 149 others

Query Match 3.0%; Score 165; DB 37; Length 903;

Best Local Similarity 32.6%; Pred. No. 2.77e-08;

Matches 115; Conservative 53; Mismatches 185; Indels 0; Gaps 0;

428 CACATTCACCTTTGCGGATCGGTCGTCGTCGTCGTCGTCGTCGTCG 487

7 CATTTCATTTTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 66

488 AATGAAGGCTATTCAGATCGATTCGATTCGATTCGATTCGATTCGAT 547

67 GATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 126

548 ACGAGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 607

127 YTNMGNAAYAAAGGATTAATTAATTAATTAATTAATTAATTAATTA 186

608 GTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 667

187 GTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 246

668 GCGGCTATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 727

247 YTNMGNAAYAAAGGATTAATTAATTAATTAATTAATTAATTAATTA 306

728 GGCATCGCATTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780

307 WSNGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 359

RESULT 3

LOCUS AF034173 2275 bp mRNA EST 30-MAR-1998

DEFINITION AF034173 Human mRNA (Tripodis and Ragousis) Homo sapiens CDNA

ACCESSION AF034173

VERSION 92707735

KEYWORDS AF034173.1 GI:2707735

SOURCE EST.

ORGANISM human.

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia;

Eutheria: Primates: Catarrhini: Hominiidae: Homo.

1 (bases 1 to 2275)

Tripodis, N. and Ragousis, J.

Generation of a transcription map in the region immediately

centromeric to human MHC across the 6p21.2-6p21.3 chromosomal

boundary.

Unpublished (1997)

On Jan 19, 1998 this sequence version replaced gi:2045115.



4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Contact: Marra M/Mouse EST Project  
WashU-HIMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108







```

E-Coli DH10B"
/db_xref="taxon:9606"
/clone="Plate-2236 Col=23 Row=D"
/clone_lib="CIR Approved Human Genomic Sperm Library D"
/sex="male"

BASE COUNT      119 a      98 c      73 g      139 t

ORIGIN

Query Match      2.8%; Score 154; DB 40; Length 429;
Best Local Similarity 33.0%; Pred. No. 6.32e-06;
Matches 34; Conservative 26; Mismatches 43; Indels 0; Gaps 0;

Db 161 AATCATTTGGCAACTATGGCTGAGATCACTGCAAGCAGCAGTGTTCCTGATC 220
      1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
Oy 1046 AAGCCTTTTNAAGARTTGGCGAAGMSVNTTNGAAGCNGAAGMGTNTTNTG 1105
      1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
Db 221 CAATGATGAGAGCTTTAGGAAATCTGCCAAGAACTTACAT 263
      1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
Oy 1106 AATATTTTGGNSNATMGNGAARAAYNSGNGCNTTACMAT 1148
      1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :

RESULT 9
LOCUS      AA063603      436 bp      mRNA      EST      24-SEP-1996
DEFINITION zeb7d08.s1 Soares_fetal_heart_NbH19W Homo sapiens cDNA clone
IMAGE:365967 3', mRNA sequence.
ACCESSION  AA063603
NID         91557570
VERSION    AA063603.1 GI:1557570
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 436)
            Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
            Holtman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
            Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
            Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
            Wilson, R.
            The WashU-Merck EST Project
            Unpublished (1995)
            On Apr 14, 1993 this sequence version replaced gi:693014.

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: polyt not found
Seq primer: 40M13 fwd. from Amersham
High quality sequence stop: 402.

FEATURES
Source
1..436
Location/Qualifiers
1..436
/organism="Homo sapiens"
/note="Organ: heart; Vector: pRT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGTGGAGGAGCGCCGACATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pRT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M. Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbH19W."
/db_xref="GDB:1282671"
/db_xref="taxon:9606"
/clone="IMAGE:365967"
/clone_lib="Soares_fetal_heart_NbH19W"

```

```

/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"

BASE COUNT      91 a      130 c      132 g      76 t      7 others

ORIGIN

Query Match      2.8%; Score 154; DB 35; Length 436;
Best Local Similarity 34.7%; Pred. No. 6.32e-06;
Matches 41; Conservative 26; Mismatches 51; Indels 0; Gaps 0;

Db 74 GAAGACTACCATGAGATGTTTGGGAACACCTCATCAATTGAGCTGCATCGAG 133
      1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
Cp 294 DATDATTTTCCNARAAATCTTTRTANSWNCNACRTCNARATTNARATGRTGNCCK 235
      1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
Db 134 CCCAGGTGTTCAGCCGCCAGTGCACAGAGCCCTGAATGACAGCCAGAGTTC 191
      1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
Cp 234 MACDATTTCTCTGTGNSNMGCRANAGCCTTCCCTGTGATMACACCATTCYTC 177
      1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :

RESULT 10
LOCUS      AA423163      455 bp      mRNA      EST      16-OCT-1997
DEFINITION ve36b02.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
IMAGE:820203 5' similar to gb:J04953 Mouse gelsolin gene, complete
cds (MUSE);, mRNA sequence.
ACCESSION  AA423163
NID         92101978
VERSION    AA423163.1 GI:2101978
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 455)
            Marra, N., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
            Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
            Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
            Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
            Waterston, R.
            The WashU-HMI Mouse EST Project
            Unpublished (1996)
            On Apr 14, 1993 this sequence version replaced gi:716974.

TITLE
JOURNAL
COMMENT

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:488483
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 165.

FEATURES
Source
1..455
Location/Qualifiers
1..455
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Organ: mammary gland; Vector: pRT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
TGTTACCAATCTGAGTGGAGGAGCGCCGACATGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pRT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."
/db_xref="taxon:10090"
/clone="IMAGE:820203"
/clone_lib="Soares mouse mammary gland NbMMG"
/sex="male"

```



```

/tissue_type="mammary gland"
/deo_stage="4 weeks"
/lab_host="DH10B"

BASE COUNT      101 a      116 c      134 g      104 t

ORIGIN

Query Match          2.8%; Score 152; DB 13; Length 455;
Best Local Similarity 39.1%; Pred. No. 1,64e-05;
Matches 34; Conservative 18; Mismatches 35; Indels 0; Gaps 0;

Db   50 AGAGGTCCTCGGCAGACTTATGCAGAGACCTGGCTACTGATGACGTATCCTCTGA 109
    1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : :
Cp   1078 ANARNSWYTCNCCRAAYLTCTRTNARAAANGCYTGNGTCKNSFRANGRTRTGNGYTGRA 1019
    1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : : 1 : : :

Db   110 CACCTGGGACCAGCTCTTGTCTGGGT 136
    1 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cp   1018 ANACNGCNACNCACYCTYTRTGNCGRT 992
    1 | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
LOCUS       W13700              459 bp      mRNA           EST           02-OCT-1997
DEFINITION  ma2ef05.r1 Soares mouse p3MMF19.5 Mus musculus cDNA clone
            IMAGE:317217 5' similar to gb:X04412 GELSOLIN PRECURSOR, PLASMA
            (HUMAN); mRNA sequence.
ACCESSION   W13700
NID         G1287738
KEYWORDS    W13700.1 GI:1287738
SOURCE      EST.
ORGANISM    house mouse.
            Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 459)
AUTHORS     Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geiselt,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
            The Mashu-HMI Mouse EST Project
            Unpublished (1996)
COMMENT     On May 8, 1995 this sequence version replaced gi:800256.
```

```

FEATURES
source
Location/Qualifiers
1..459
/organism="Mus musculus"
/note="vector: pRTT3 (pharmacia) with a modified
polylinker; site_1: Not I; site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGGCGCCGCATTCTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pRTT3 vector
(Pharmacia). Library went through one round of
normalization to a Cot =5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Mayne State University)."
/db_xref="taxon:10090"
/clone="IMAGE:317217"
/clone_lib="Soares mouse p3NMF19.5"
/dex_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"

```

BASE COUNT	104	a	123	c	130	g	102	t
ORIGIN								
Query Match			2.8%	Score 152;	DB 33;	Length 459;		
Best Local Similarity			39.1%	Pred. No.	1,64e-05;			
Matches	34;	Conservative	18;	Mismatches	35;	Indels	0;	Gaps 0.
Db	99	AGAGGTTCCTGGCGAGGCTTATGACGAAGAACCCTGGCTACTGATGACGCATCCTCTCGA	158					
	:	::	:	: :	: :	: :	: :	: :
Cp	1078	ANARNSWTCNGCAAAATCRTNRRARRANGCYTGNCINCSMWANAAGTRTNGGYGRA	1019					
Db	159	CACCTGGGACCCAGGTCCTTTGTCTGGGT	185					
			:	:	:	:	:	:
Cp	1018	ANACGCCMACNACYCTTTRTNGNGRT	992					

LOCUS	12	RESULT
DEFINITION	468 bp DNA	GSS 15-DEC-1998
ACCESSION	U00000.1	
VERSION	U00000.1	
KEYWORDS	Sequence	
SOURCE	Human	
ORGANISM	Homo sapiens	
REFERENCE	Miller, R., et al. 1998. The Human Genome Project: A Review of the Progress and the Challenges Ahead. <i>Science</i> 281: 205-211.	
AUTHORS	Shake, R., Schmidt, S., Trautwein, R., and Hood, L.E.	
TITLE	Construction of a Characterized Clone Resource for Genomic Sequencing	
JOURNAL	Unpublished (1998)	
COMMENT		

Contact: Malaitra GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel.: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 2190 Row: E Column: 6  
Class: BAC ends  
High quality sequence stop: 466.

[illegible]



```
KEYWORDS      EST . house mouse.
SOURCE        Mus musculus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 505)
AUTHORS       Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T...
              Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,T., Morris,M.,
              Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
              Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
              Waterston,R.
TITLE         The MashU-HMMI Mouse EST Project
JOURNAL       Unpublished (1996)
COMMENT       On Sep 12, 1996 this sequence version replaced gi:1393053.

Contact: Marra M/Mouse EST Project
MashU-HMMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel.: 314 286 1800
Fax: 314 286 1810
Email: mouseset@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (infoimage.lnl.gov) for further information.
MG1:33310
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 356.
Location/Qualifiers
    1..505
    /organism="Mus musculus"
    /strain="C57BL/6 x CBA"
    /note="Organ: Lung; Vector: pBluescript SK-; Site_1:
    EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
    Oligo dt. 6-8 month old female lung and 1.5 year old male
    lung were source of mRNA. Average insert size: 1.5 kb;
    Uni_ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGACGAGCA
    3' -3' adaptor sequence: 5' CTCGAATTCTTTTCTTTTTTTT 3'"
    /db_xref="taxon:10090"
    /clone_id="Stratagene mouse lung 937302"
    /seq="female"
    /issue_type="lung"
    /dev_stage="6-8 month old"
    /lab_host="SOLR (Kanamycin resistant)"
BASE COUNT   113 a 133 c 147 g 111 t 1 others
ORIGIN
Query Match          2 8% Score 152; DB 36; Length 505;
Best Local Similarity 39.1%; Pred. No. 1,64e+05;
Matches 34; Conservative 18; Mismatches 35; Indels 0; Gaps 0;

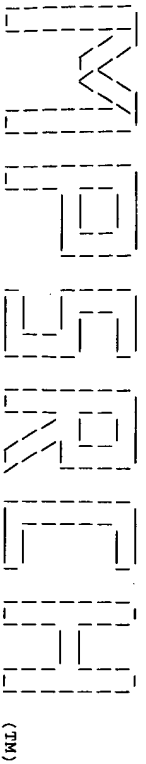
Db 100 AGAGGTCCCTGCAGCGACTTAGCAGAAGACCCTCATCTGATGCATCATCTCTCGTA 159
| : :: ||::||: |::|| |::|| |::| |::| |::| |::|
Cc 1078 ANARSWTCNGRCRAAYCRITNARRANGCYTGTCGNCKNSMRANGRTRONGYTGRA 1019
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 160 CACCTGGACCAAGCTCTTTGTCGTGGCT 186
| | | | | | | | | | | | | | | | | | | | | | | | | |
Cc 1018 ANACNGCNACNACYCTTPTGNGRT 992

RESULT 15 AA839460 515 bp mRNA EST 27-FEB-1998
LOCUS wv50c12.r1 Soares mouse mammary gland NbMG Mus musculus cdNA clone
DEFINITION IMAGE:1247254 5' similar to gb:j04953 Mouse gelsolin gene, complete
cds (mouus);, mRNA sequence.
ACCESSION AA839460
NID 92915555
VERSION AA839460.1 GI:2915555
KEYWORDS EST .
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```









Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Merch\_nrp n.a. - n.a. Smith-Waterman search, using a protein database  
which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Nov 24 01:53:04 1999; Maspar time 95.56 Seconds  
881.244 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-3  
Description: (1-660) from US09103287.seq  
Perfect Score: 3300  
N.A. Sequence: 1 ATTAAAGATTCGATGACA.....CGCTTTATATGCTTTATA 660  
Comp: TAAATTTCTACGCTACTGT.....CGCAAAATTAACAATAT

Scoring table: TABLE bkttranslated2  
Gap 30

Match STD : Dbase 0; Query 0

Searched: 170751 seqs, 63799824 bases x 2

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database:

a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 55.052; Variance 302.910; scale 0.182

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	2557	77.5	215	39	W89199	Partial sequence of t
2	2557	77.5	437	39	W87771	UDP-N-acetylmuramate: 1.45e-178
3	764	23.2	422	33	W55120	Streptococcus pneumoniae: 9.55e-42
4	534	16.2	46	34	W7686	Staphylococcus aureus 5.57e-25
5	182	5.5	335	33	W55117	Streptococcus pneumoniae 2.57e-01
6	182	5.5	450	37	W68551	S. pneumoniae Murd pr 2.57e-01
7	182	5.5	450	26	W29454	Streptococcus pneumoniae 4.39e-01
8	178	5.4	286	22	W24585	H. pylori cytoplasmic 4.39e-01
9	178	5.4	286	22	W20102	H. pylori cytoplasmic 4.39e-01
10	178	5.4	455	22	W20606	H. pylori cytoplasmic 4.39e-01
11	145	4.4	37	36	W7378	Lytic peptide with a1 3.02e+01
12	143	4.3	993	17	R95268	Pre-nisin modification 3.86e+01

13	140	4.2	40	15	R84924	Alpha-helix-forming o
14	137	4.2	540	22	W20605	H. pylori cytoplasmic
15	137	4.2	854	10	R51250	Flv PPR envelope prot
16	134	4.1	244	3	P50304	Storage-specific, lat
17	136	4.1	354	11	R58924	Alpha subunit of huma
18	134	4.1	552	20	W10070	Murine long form CSF
19	134	4.1	552	24	W22615	Murine 4 kb colony st
20	134	4.1	552	1	P80766	Deduced sequence for
21	134	4.1	552	24	W22615	Murine 2 kb colony st
22	134	4.1	552	27	W36146	Mouse short form of c
23	134	4.1	552	26	W35753	Murine colony stimula
24	134	4.1	552	26	W35752	Murine colony stimula
25	134	4.1	552	20	W10071	Murine short form CSF
26	134	4.1	552	27	W36145	Mouse long form of co
27	134	4.1	552	1	P80390	Deduced sequence for
28	134	4.1	224	24	W26357	Human LDL receptor an
29	131	4.0	30	30	W55873	HIV p17 gag protein a
30	131	4.0	176	1	P90414	Plasmodium falciparum
31	131	4.0	308	26	W34322	Product of restrictio
32	131	4.0	431	27	W34528	Protein encoded by pl
33	131	4.0	431	18	R98004	Oestrogen-regulated p
34	132	4.0	464	39	W86321	Kidney injury associa
35	133	4.0	530	24	W26428	Swinepox virus HindII
36	130	3.9	75	24	W25779	Peptide carrier #2 fo
37	130	3.9	269	19	W02715	G-protein coupled odo
38	130	3.9	269	16	R48743	G-protein coupled odo
39	130	3.9	395	18	W03474	Mouse SVR-related pro
40	130	3.9	401	6	R28802	Genetic BT toxin sequ
41	130	3.9	401	6	R32848	Genetic BT toxin sequ
42	130	3.9	568	12	R66452	Serine-rich AF-9 prot
43	130	3.9	716	6	R28801	Genetic BT toxin sequ
44	130	3.9	716	6	R32847	Genetic BT toxin sequ
45	129	3.9	2109	19	W04540	Vesiculovirus large (

## ALIGNMENTS

RESULT 1  
ID: W89199 standard; Protein: 215 AA.  
AC W89199;  
DE Partial sequence of the Murc polypeptide.  
KW Murc gene; UDP-N-acetylmuramate:L-alanine ligase; Murc polypeptide;  
KW bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
KW immunogen; drug; genetic immunisation.  
OS Staphylococcus aureus.  
PN EP-889123-A2.  
PD 07-JAN-1999.  
PE 26-JUN-1998; 305064.  
PR 03-JUL-1997; US-052720.  
PA (SMK ) SMITHKLINE BEECHAM CORP.  
PI (SMK ) SMITHKLINE BEECHAM PLC.  
PI Burnham MKR, Wallis NG;  
DR WPI: 99-062655/06.  
DR N-PSDB: V80065.  
PT New isolated Murc polypeptide from Staphylococcus aureus and related  
PT nucleic acid, useful in diagnosis, treatment and prevention of  
PT bacterial infections  
PS Claim 1; Page 5; 39pp; English.  
CC The invention relates to a UDP-N-acetylmuramate:L-alanine ligase  
CC (Murc polypeptide) encoded by the S. aureus Murc gene. Host cells  
CC containing an expression system comprising the Murc gene can be used for  
CC the recombinant production of the polypeptide. Agonists or the Murc  
CC polypeptide are used to treat conditions requiring increased activity or  
CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or  
CC competitive polypeptide are useful for inhibiting the polypeptide e.g.  
CC bacterial (especially S. aureus) infections. They are also useful against  
CC Helicobacter pylori infections and related cancers, ulcers and gastritis.  
CC The antibacterial agents are useful to treat in-dwelling devices for  
CC infection prevention or generally as wound treatments to prevent adhesion  
CC of bacteria to matrix proteins. The Murc polypeptide is also useful for  
CC diagnosing or prognosing a (susceptibility to) disease, for raising  
CC antibodies; to identify modulators or specific receptors; in rational







CC antibodies in standard immunoassays, especially for diagnosing or  
CC monitoring infections. Antibodies which bind the protein are used to  
CC detect corresponding antigens, to purify the protein and for passive  
CC immunisation (optionally coupled to a toxin). Vaccines are administered  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
CC (especially 10-300) mu g/ml per dose.

Sequence 422.AA:

Query Match 23.2%; Score 764; DB 33; Length 422;  
Best Local Similarity 40.4%; Pred. No. 9,556-42;  
Matches 245; Conservative 106; Mismatches 249; Indels 6; Gaps 4

Db Dc  
623 d f v a s d l l r s i t g s t f t v n h f  
15 ATGACATTATTCGACCAATATTTCAAAATACGGATTAAGGATACGCTGTTATGATG 74  
D I Y A Q I F Q I T D K G T A V D V Y V  
Db Dc  
683 t r g q n l g q f i p t f g t r h n i m n  
75 TGGATGGATGCTTTATATCATCTACTCGTCCGTCCACATATGATGACCATTCAGATTAA 134  
D G E F Y D H F L S P Q Y G D H T V L N  
Db Dc  
743 a t a v i g l l y t a g f d l n l v r e  
135 ATGCTATTCGCTATTCGATTGTTATTTAGGAAACCTGATGTTCAATATTTAAG 194  
A L A V I A I S Y L E K L D V T N I K E  
Db Dc  
803 h l k t f a g v k r f t f t e k i v n d t  
195 AAGCATTTGAACGCTTGGTGTAAACGCTGTTCAATGAACCTCAATTGCAATTC 254  
A L E T F G G V K R R F N E T T I A N Q  
Db Dc  
863 c n g n t h a g y a y l y g n c a y c n a c n g a r a t h a g n c n l n g a g n c  
255 AAGTATTTGATGATTTATGACACACACCTCCAGAGAAATAGTCCTCAATTTGACACAG 314  
I V D D Y A H H P R E I S A T I D T A  
Db Dc  
923 r q k y p s k e i v a v a v f q p h t f t r  
315 CACGAAAGAAATATCCACATTAAGAGTGTTCGAGTATTCACACACACCTTCTCTA 374  
R K K Y P H K E V V A V F Q P H T F S R  
Db Dc  
983 t i a l l d d f a h a l n q a d a v y l  
375 GAACACAGCAATTTTAATGAATTCGAGAAAGTTAAGTAAGCAACATGCTATTC 434  
T Q A Y F L N E F A E S L S K A D R V F L  
Db Dc  
1043 t n g c a r a t h l a y g n s n g m n g r t n g a y c a g g a y t n a a r t n g a r t  
435 TATGGAATTTGTTGGATCAATTCAGAGAAAT-AC-TCGGCAATTATGATACAGATT 491  
C E I F G S I R E N X X G A L T I Q D L  
Db Dc  
1103 t n g c n a a y a a r a t h a y a a r a c a r g n a t h a c n g t n a a y t n s c n y n l  
492 TAAATGATTAATATGA--AGTGATC-GTTAATTAAAGAGATTCATATGATATG 548  
I D K I X X V H X L I N E D S I N V L E  
Db Dc  
1163 t n g a y c a y a a y c n g t n l a y g n t l y a c g g n g a g g d i q t y e y  
549 AACATCTGATTAATCCCTGTTTATTTATGAGCGGCGAGGATATTCAAAATTACAA 608  
Q F D N A V V L F M G A G D I Q K L Q N  
Db S



Dt 1223 aysnt 1228  
QY 609 ATGCAT 614  
Qt A

RESULT 4  
ID W77686 standard; Protein; 46 AA.  
AC W77686:  
DE 30-OCT-1998 (first entry)  
KW Staphylococcus aureus protein of unknown function.  
KW Staphylococcus aureus protein: immune response induction; eye infection;  
antibody production; T-cell immune response; gastrointestinal infection;  
respiratory infection; inhibitor; bacterial infection; cardiac infection;  
central nervous system; kidney infection; urinary tract infection;  
antimicrobial compound identification; broad spectrum antibiotic;  
therapy.  
KW Staphylococcus aureus.  
OS EP-841394-A2.  
PN 13-MAY-1998.  
PD 24-SEP-1997: 307485.  
PR 24-SEP-1996: US-027032.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PI Black MT, Burnham MKR, Hodgson JM, Knowles DJC,  
P1 Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,  
P1 Ward JM:  
WPI: 98-252940/23.  
DR N-PSDB: V33479.  
PT New nucleic acid sequences from Staphylococcus aureus WCHU29 -  
PT useful in vaccines and for treatment of bacterial infections of e.g.  
PT respiratory tract and central nervous system  
PS Claim 11: Page 329: 390pp: English.  
CC This sequence represents a Staphylococcus aureus protein of unknown  
CC function, and is encoded by a DNA sequence of the invention.  
CC The DNA sequences were isolated from Staphylococcus aureus WCHU29  
CC (NCIMB 40771). Host cells containing the DNA sequences are used to  
CC produce polypeptides or fragments. The proteins are used in the treatment  
CC of disease, for inducing an immune response by administering them, to  
CC produce antibody and/or T-cell immune response. Antagonists of the  
CC proteins are used for the inhibition of bacterial polypeptides.  
CC Conditions which may be treated include bacterial infections, especially  
CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,  
CC urinary tract, skin, bones and joints. The proteins can also be used to  
CC identify antimicrobial compounds which are broad spectrum antibiotics,  
CC especially useful in the treatment of H. pylori infection.  
SQ Sequence 46 AA.

Query Match 16.2%; Score 534; DB 34; Length 46;  
Best Local Similarity 61.6%; Pred. No. 5.57e-25;  
Matches 85; Conservative 32; Mismatches 20; Indels 1; Gaps 1;

Dt v h r s l m k i l i n v l e g f d n a v  
Dc 1 g t c a y m g n w n y t a r g a a r t h y t n a t h a y g n y n g a r c a t t y g a y a y g c n l n 60  
QY 510 G T G C A T C C G T T A T T A T A A G A T T C T - A T T A T G T A T T A G A A C A T T G A T T C T C T T 568  
Qt V H R \* L M K I N V L A E Q F D N A V

Dt v l f m g a g d i g k l q n a y l d k l  
Dc 61 g t n y n t l y a t g g n g n g n g a y t h a r a a r y t n c a a r y g c n t a y t n g a y a a r y t n 120  
QY 569 G T T T A T T A T G T G C A G C A G T A T T C A A A A A T T A C A A A A T T A T T A T T A T T A A A T T A 628  
Qt V L F M G A G D I Q K L Q N A Y L D K L

Dt g m k n a f  
Dc 121 g g n a t g a a r a a y g c n t y 138  
QY 629 G G C A T G A A A A A T G C C T T 646  
Qt G M K N A F

RESULT 5  
ID W55117 standard; Protein; 335 AA.

AC W55117;  
DT 02-OCT-1998 (first entry)  
DE Streptococcus pneumoniae SP0067 protein.  
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
KW detection; pneumonia; otitis media; meningitis.  
OS Streptococcus pneumoniae.  
FH Key Location/Qualifiers  
FT Misc.difference 1  
FT /note= "encoded by T"  
PN W09818930-A2.  
PD 07-MAY-1998.  
PF 30-OCT-1997: U19422.  
PR 31-OCT-1996: US-029960.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Choi GH, Bromocky J A, Johnson LS, Kunsch CA;  
DR WPI: 98-272224/24.  
DR N-PSDB: V27378.  
PT Nucleic acid encoding antigenic peptide(s) from Streptococcus  
PT pneumoniae - or their epitope-containing fragments, useful in  
PT protective or therapeutic vaccines, and for diagnosis  
PS Claim 11: Page:72: 118pp: English.  
CC The present sequence represents a protein from Streptococcus pneumoniae.  
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
CC can be useful in vaccines for inducing protective antibodies against  
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
CC are used to detect Streptococcus infection (by usual hybridisation or  
CC amplification methods), also for isolating Streptococcus genes or their  
CC allelic variants. The protein can be used similarly to detect specific  
CC antibodies in standard immunoassays, especially for diagnosing or  
CC monitoring infections. Antibodies which bind the protein are used to  
CC detect corresponding antigens, to purify the protein and for passive  
CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
CC (especially 10-300) mu g/ml per dose.  
SQ Sequence 335 AA.

Query Match 5.5%; Score 182; DB 33; Length 335;  
Best Local Similarity 41.8%; Pred. No. 2.57e-01;  
Matches 46; Conservative 18; Mismatches 46; Indels 0; Gaps 0;

Dt h n v e n a l a t i a v a k l r d v d n  
Dc 487 c a y a a y t n g a r g a c n y t n g c n a t h y c n g t n g a a r y t n m g n g a y t n g a y a y 546  
QY 122 C A T C A G T T T A A T T A A T T A A G C A T T A G C G T A T T A T T A G A A C G C T A G A G T T 181  
Qt H T V L N A L A V I A I S Y L E K L D V

Dt g t l k e t l s a f g g v k h r  
Dc 547 c a r c a n a t h a r g a r a c n y t n s n g c n t t y g n g n g n t n a r c a y m g n y t 596  
QY 182 A C A A T A T T T A A G A A C A T T A G A A C G T T G C G T G T T A A C G T C T T 231  
Qt T N I K E A L E T F G G V K R R

RESULT 6  
ID W68551 standard; Protein; 450 AA.  
AC W68551:  
DE 25-JAN-1999 (first entry)  
KW S. pneumoniae MURD protein.  
KW Biosynthesis; recombinant; antibacterial; bacterial cell wall;  
uridine-diphosphate-N-acetylglucosaminyl-L-alanyl-D-isoglucamate ligase;  
KW drug design; resistance; microorganism.  
OS Streptococcus pneumoniae.  
PN US5834270-A.  
PD 10-NOV-1998.  
PF 14-APR-1997: 843309.  
PR 18-JUN-1996: US-665435.  
PR 14-APR-1997: US-843309.  
PA (Etil.) LILLY & CO. ELI.  
PI Hoskins JA, Peery RB, Skatrud PL, Wu CE;  
DR WPI: 99-008720/01.  
DR N-PSDB: V33970.  
PT Streptococcus pneumoniae mur D biosynthetic gene - which encodes







OY           || ||:|:|:|   ||:|:|   ||   ||:|:|   ||:|:|   ||:|:|   ||:|:|  
       275 GCACACCATCCAGAGAATTAGTCTACAA TTGACACGACGAA 320  
 qt       A H H P R E I S A T I D T A R

RESULT	9
ID	W20102 standard; Protein; 286 AA

Query Match	5.48;	Score 178;	DB 22;	Length 286;
Best Local Similarity	37.38;	Pred. No. 4.39e-01;		
Matches	62;	Conservative 33;	Mismatches 69;	Indels 2;
			Gaps	2;

[illegible]

QY 275 GCACACCATCCAGAGAAATTGCTACATTTGACACAGCAGAA 320  
 A H H . P R E I S A T I D T A R  
 Q\*

RESULT	10
ID	W20606 standard; protein; 455 AA

Query Match	5.48;	Score 178;	DB 22;	Length 455;
Best Local Similarity	38.08;	Pred. No. 4.39e-01;		
Matches	63;	Conservative	34;	Mismatches 67;
			Indels	2;
			Gaps	2

Dd s a l d e l n l e e l r n n l l n f k g  
Dt 860 twnsgcnylngaygarycnaaytngargaratlmgnaaayaytynytnaaytlyarg 911

OY 156 TAGTATTATAGAGAAGCTACGTTCAAAATTTTAAGAGCGATTGAAGAAGTGTCGTG 21:  
Qt S Y L E K L D V T N I K E A L E T F G G

Dd | k k r t f d i l q k n d l i l i d g y  
920 gnathaaaraangtcttgay-athytcacaaaacaagaygattynathyttnathgayray 978:  
OY | :  
Qt 216 GGCTAAGCCTGCCTTCATCAACTGAACCTGCCA-ATCAAGCTTCCTTAAGTATATTAT 27:  
Dd v k r r e f c n k l o l o x o v i v d d d y

```
Dbb      a h h p t e i g t c t l k s a r  
Dc    979 gcacaycaycnagarttggnaacnacylnaarwsgncmnga 102  
        ||||| : | : | : | : | : | : | : |  
Qy   275 GCACACCCTCCAGAAATTAAGTGTCAATTGCACGCGAA 320  
        A H H P R E I S A T I D T A R
```

RESULT	11
ID	w77378 standard; peptide; 37 AA
AC	w77378;
DT	14-DEC-1998 (first entry)















CC LENGTH: 450 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 450 AA; 48579 MW; 1010380 CN;

Query Match  
Best Local Similarity 41.8%; Score 182; DB 2; Length 450;  
Pred. No. 2.59e-01;  
Matches 46; Conservative 18; Mismatches 46; Indels 0; Gaps 0;

DB H N V E N A L A T I A V A K L R D V D N  
Dc 832 CAYAYGNGARAAVGYNGCNACNATHGNGNGNARAYTMNGAYGNGAYAY 891  
QY 122 CATACAGTTTAAATGCACTAGCTGTAATGTTAGTATTTAGAGAGCTAGATGTT 181  
Qt H T V L N A L A V I A I S Y L E K L D V

Db O T I K E T L S A F G G V K H R  
Dc 892 CARACNATHARAGACNYTMNSGNTTYGNGNGNARAYTMNGAYGNGAYAY 941  
QY 182 ACAATATTAAAGAGCATTTGAACGTTGCTGCTTAACGCTGTTT 231  
Qt T N I K E A L E T F G G V K R R

RESULT 2  
ID US-08-665-435A-2 STANDARD: PRT: 450 AA.  
XX xxxxxx

Sequence 2, Application US/08665435A  
Patent No. 5681694  
GENERAL INFORMATION:  
APPLICANT: Skatrud, Paul  
APPLICANT: Peery, Robert  
APPLICANT: Hoskins, JoAnn  
APPLICANT: Wu, Chyun-Yeh Earnest  
TITLE OF INVENTION: Blosynthetic Gene Mumd of Streptococcus  
TITLE OF INVENTION: pneumoniae  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
City: Indianapolis  
STATE: Indiana  
COUNTRY: US  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/665,435A  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Webster, Thomas D  
REGISTRATION NUMBER: 39,872  
REFERENCE/DOCKET NUMBER: X-9900  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3334  
TELEFAX: 317-276-3861  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 450 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 450 AA; 48579 MW; 1010380 CN;

Query Match  
Best Local Similarity 41.8%; Score 182; DB 1; Length 450;  
Pred. No. 2.59e-01;  
Matches 46; Conservative 18; Mismatches 46; Indels 0; Gaps 0;

DB H N V E N A L A T I A V A K L R D V D N  
Dc 832 CAYAYGNGARAAVGYNGCNACNATHGNGNGNARAYTMNGAYGNGAYAY 891  
QY 122 CATACAGTTTAAATGCACTAGCTGTAATGTTAGTATTTAGAGAGCTAGATGTT 181  
Qt H T V L N A L A V I A I S Y L E K L D V

Db O T I K E T L S A F G G V K H R  
Dc 892 CARACNATHARAGACNYTMNSGNTTYGNGNGNARAYTMNGAYGNGAYAY 941  
QY 182 ACAATATTAAAGAGCATTTGAACGTTGCTGCTTAACGCTGTTT 231  
Qt T N I K E A L E T F G G V K R R

RESULT 3  
ID US-08-934-481-4 STANDARD: PRT: 267 AA.  
XX xxxxxx

Sequence 4, Application US/08934481  
Patent No. 5928045  
GENERAL INFORMATION:  
APPLICANT: Wallis, Nichola G.  
APPLICANT: Fuenyo, Joanna L.  
APPLICANT: Lonetto, Michael A.  
TITLE OF INVENTION: NOVEL Mumd  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
City: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/934,481  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dickinson, Todd Q  
REGISTRATION NUMBER: 28,354  
REFERENCE/DOCKET NUMBER: GM10070A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2252  
TELEFAX: 215-994-2222  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 267 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE 267 AA; 29079 MW; 348410 CN;

Query Match  
Best Local Similarity 40.9%; Score 177; DB 2; Length 267;  
Pred. No. 4.83e-01;  
Matches 45; Conservative 18; Mismatches 47; Indels 0; Gaps 0;



Db H N V E N A L A T I A V A K L R G V D N  
Dc 283 CAYAAAGTNGARAAAGTNTGNCNACNATHGCGTNGCNAARTYNNMGNGNGAYAAV 342  
Oy 122 CATACGTTTAAATGATAGTGTAAATTCATAGTATTTAGAGAGATGTT 181  
Qt H T V L N A L A V I A I S Y L E K L D V

Db O T I K E T L S A F G G V K H R  
Dc 343 CACACNATHAARGARACNTYNNMGNGTNAARCAVNGNTY 392  
Oy 182 ACAATATTTAAAGACATAGAAAGCTTGGTGTGTTAAAGCTGTT 231  
Qt T N I K E A L E T F G G V K R R

RESULT 4  
ID US-08-934-481-2 STANDARD; PRT; 446 AA.  
xx xxxxxx

Sequence 2, Application US/08934481

Sequence 2, Application US/08934481  
Patent No. 5929045  
GENERAL INFORMATION:  
APPLICANT: Wallis, Nicholas G.  
APPLICANT: Fuenyo, Joanna L.  
APPLICANT: Lonetto, Michael A.  
TITLE OF INVENTION: NOVEL MURD  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/934.481  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dickinson, Todd O  
REGISTRATION NUMBER: 28,354  
REFERENCE/DOCKET NUMBER: G410070A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2252  
TELEFAX: 215-994-2222  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 446 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE 446 AA: 47982 MW: 998066 CN;

Query Match 5.4%; Score 177; DB 2; Length 446;

Best Local Similarity 40.9%; Pred. No. 4.83e-01;  
Matches 45; Conservative 18; Mismatches 47; Indels 0; Gaps 0;

Db H N V E N A L A T I A V A K L R G V D N  
Dc 820 CAYAAAGTNGARAAAGTNTGNCNACNATHGCGTNGCNAARTYNNMGNGNGAYAAV 879  
Oy 122 CATACGTTTAAATGATAGTGTAAATTCATAGTATTTAGAGAGATGTT 181

Qt H T V L N A L A V I A I S Y L E K L D V  
Db O T I K E T L S A F G G V K H R  
Dc 880 CACACNATHAARGARACNTYNNMGNGTNAARCAVNGNT 929  
Oy 182 ACAATATTTAAAGACATAGAAAGCTTGGTGTGTTAAAGCTGTT 231  
Qt T N I K E A L E T F G G V K R R

RESULT 5  
ID US-08-687-551-6 STANDARD; PRT; 40 AA.  
xx xxxxxx

Sequence 6, Application US/08687551

Sequence 6, Application US/08687551  
Patent No. 5856435  
GENERAL INFORMATION:  
APPLICANT: BAZILE, Didier  
APPLICANT: EMILE, Carole  
APPLICANT: HELENE, Claude  
APPLICANT: SPENLEHAUER, Gilles  
TITLE OF INVENTION: NUCLEIC ACID-CONTAINING COMPOSITION, ITS  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Rd. 3C43  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19426  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,551  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 94/01381  
FILING DATE: 08-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR95/00098  
FILING DATE: 27-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith Ph.D., Julie K.  
REGISTRATION NUMBER: 38,619  
REFERENCE/DOCKET NUMBER: ST94007-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610)454-3839  
TELEFAX: (610)454-3808  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 40 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE 40 AA: 4845 MW: 9430 CN;

Query Match 4.2%; Score 140; DB 2; Length 40;

Best Local Similarity 27.7%; Pred. No. 4.01e-01;  
Matches 33; Conservative 31; Mismatches 35; Indels 0; Gaps 0;

Db L K K L L K K L L K K L L K L L K K L  
Dc 1 YNNAARARYTYNNAARARYTYNNAARARYTYNNAARARYTYNNAARARYTN 60  
Oy 1 YNNAARARYTYNNAARARYTYNNAARARYTYNNAARARYTYNNAARARYTN 60



DT



[illegible]

```
CC      TITLE OF INVENTION:    NUCLEIC ACID FRAGMENTS DERIVED FROM THE
CC      TITLE OF INVENTION:    HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
CC      TITLE OF INVENTION:    APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
CC      TITLE OF INVENTION:    MATERNOFETAL TRANSMISSION OF HIV-1
CC      NUMBER OF SEQUENCES:   130
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
CC      STREET: 1800 M Street, N.W.
CC      CITY: Washington
CC      STATE: D.C.
CC      COUNTRY: USA
CC      ZIP: 20036-5869
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: Patentin Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/649,991
CC      FILING DATE: 17-MAY-1996
CC      CLASSIFICATION: 435
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: FR 9505914
CC      FILING DATE: 18-MAY-1995
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Adler, Reid G.
CC      REGISTRATION NUMBER: 30,988
CC      REFERENCE/DOCKET NUMBER: ORES-5003
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: 202-467-7000
CC      TELEFAX: 202-467-7176
CC      INFORMATION FOR SEQ ID NO: 108:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 132 amino acids
CC      TYPE: amino acid
CC      STRANDEDNESS:
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: peptide
CC      SEQUENCE 132 AA; 14860 MW; 86327 CN;
SQ
Query Match          4.2%; Score 137; DB 2; Length 132;
Best Local Similarity 51.18; Pred. No.5.64e+01;
Matches 23; Conservative 12; Mismatches 10; Indels 0; Gaps 0;
Dd      K I D V K D T K E A L E K I
Dt      269 AAAAATGAGTGNARGAACNARGARGCANTGARAAATHG 313
Qt      168 AGAACCTGCATTTCAATATTAAAGAAGCATTTGAACAAGTTTG 212
       K L D V T N I K E A L E T F
RESULT_10
ID      US-08-649-991-86 STANDARD; PRT; 132 AA.
XX      xxxxxx
XX
DT
DE      Sequence 86, Application US/08649991
CX      Sequence 86, Application US/08649991
CC      Patent No. 5919462
CC      GENERAL INFORMATION:
CC      APPLICANT: Narwa, Remy
CC      APPLICANT: Roques, Pierre
CC      TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
CC      TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
CC      TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
CC      TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
CC      NUMBER OF SEQUENCES: 130
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
CC      STREET: 1800 M Street, N.W.
```







CC INFORMATION FOR SEQ ID

CC LENGTH: 552 amino acids











/db\_xref="PID:g2642659"  
 /db\_xref="GI:2642659"  
 /translation="MTHYHEVINGISGMSLSIAQIMHDLGHEVQSDIENYFEVALR  
 NKGIKIFEDANNIKEDMNVVLOGNAPASHEERKAROMKIDVYSYNDFOGQIDDOYT  
 SVAYTGAGKSTGTGLSHVNGDKTSELDGSDGKIDVYDFAEAGEYRHRFIS  
 YKPDYATMNTIDPDHPDYFKDINDVPADFOMANNVKGITAMDDHLKIEADVFI  
 YYGFKRSDDIYAGNIQITDKGTAFDYVDEEFDHLSPOYGHVLYNLAVIAISY  
 LEKLDVNIKEALETFEGVKRRENETIANOVIVDDAHHPREISATETARKKYPKR  
 EVVAEPQPTFSRTQAFLENEFSLRADRFADICEIGISIRENTGALTIDLDIKIEG  
 ASLINESINVLEBOFDNAVYLFKAGCQIOKLQNLNLYDKLGMKNF"

BASE COUNT 467 a 163 c 250 g 434 t  
 ORIGIN

Query Match 96.2%; Score 635; DB 17; Length 1314;  
 Best Local Similarity 98.9%; Pred. No. 0.00e+00;  
 Matches 642; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 666 ATTAAAGATTCGGATGATGATTAATGCTCAAAATTTCAAAATTAAGGTAAGTACGTC 725  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 1 ATTAAAGATTCGGATGATGATTAATGCTCAAAATTTCAAAATTAAGGTAAGTACGTC 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 726 TTTTGATGTATGTGATGGTGAATTTATGATCACTTCCTGCTCCACAATATGATGA 785  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 61 TGTGATGTATGTGATGGTGAATTTATGATCACTTCCTGCTCCACAATATGATGA 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 786 CCATACAGATTAAATGATGATGCTGTAATGCGATTAATTTAGAGAAGCTAGATGT 845  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 121 CCATACAGATTAAATGATGATGCTGTAATGCGATTAATTTAGAGAAGCTAGATGT 180  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 846 TACAATTTAAAGAGATTAAGAACTTTGGTGTGTTAAAGCTGTTCAATGAAC 905  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 181 TACAATTTAAAGAGATTAAGAACTTTGGTGTGTTAAAGCTGTTCAATGAAC 240  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 906 TACATTTGCAATCAAGTTATTTAGATGATGATGATGATGATGATGATGATGATGATG 965  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 241 TACATTTGCAATCAAGTTATTTAGATGATGATGATGATGATGATGATGATGATGATG 300  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 966 TACATTTGCAATCAAGTTATTTAGATGATGATGATGATGATGATGATGATGATGATG 1025  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 301 TACATTTGCAATCAAGTTATTTAGATGATGATGATGATGATGATGATGATGATGATG 360  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1026 ACACACTTCTCTGAACACAGCAATTTTAAATGAATTTGACAGAAATTTAAAGTAAAGC 1085  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 361 ACACACTTCTCTGAACACAGCAATTTTAAATGAATTTGACAGAAATTTAAAGTAAAGC 420  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1086 AGATCGTATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1145  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 421 AGATCGTATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1146 GATCAGAGATTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1205  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 481 GATCAGAGATTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1206 TGTATTGAACAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1265  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 541 TGTATTGAACAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1266 ATTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1314  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 601 ATTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 649  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 2 BACUNAM 2320 bp DNA BCT 26-MAR-1997

LOCUS Bacillus subtilis UDP-N-acetylmutamate-lysine ligase gene, partial  
 DEFINITION cds and 3 ORF's.

ACCESSION U31845  
 NID 9556013  
 VERSION L31845.1 GI:556013

KEYWORDS UDP-N-acetylmutamate-L-alanine ligase.  
 SOURCE Bacillus subtilis.  
 ORGANISM Bacillus subtilis

Eubacteria; Firmicutes; Low G+C gram-positive bacteria;

REFERENCE 1 (bases 1 to 2320)  
 AUTHORS Varon,D., Brody,M.S. and Price,C.W.  
 TITLE Bacillus subtilis operon under the dual control of the general  
 stress transcription factor sigma B and the sporulation  
 transcription factor sigma H  
 JOURNAL Mol. Microbiol. 20 (2), 339-350 (1996)  
 FEATURES 96310371  
 SOURCE location/Qualifiers  
 1..2320  
 /organism="Bacillus subtilis"  
 /strain="168"  
 /sub\_species="Marburg"  
 /db\_xref="taxon:1423"  
 <1..1049  
 /note="identification based on sequence similarity to E.  
 coli enzyme"

CDS

/codon\_start=3  
 /transl\_table=11  
 /product="UDP-N-acetyl mutamate-alanine ligase"  
 /protein\_id="AAB40043.1"  
 /db\_xref="PID:g556014"  
 /db\_xref="GI:556014"  
 /translation="IPVIRHKEFLGDYMKFTSVAVTGAHKTSTGLLAHVIONAKP  
 TSFLIDGOGGENSEEVEFACEYRHFLLSYOPDYAMTNDPDHPYFSSIDPVE  
 DAFQENALGVNKGIIACGDDEHLPIKHPVYVYGTGENDEQAKINVKSTGTFD  
 VFRNPFYDFEYIPAYGHNVLNLSLAVIALCHEEIDSSIGIKRAKSPGVKRENEK  
 OGDVLDIDYAHFPEITKVTLEAKOKYDPREIVAPPHFTYRQQLDEABESL  
 GACVYLCDFGSAFENAGKLTIGDLOGKIHNNKLTIEDDTSVLKAHNAVLIFKAG  
 DIQKIRATEYNVA"  
 1048..1079  
 /note="sigma-H consensus"  
 1101..1109  
 /note="sigma-H consensus"  
 1123..1127  
 /note="sigma-H consensus"  
 1136  
 /note="sigma-H dependent 5' end"  
 1151..1158  
 /note="sigma-B consensus"  
 1172..1177  
 /note="sigma-B consensus"  
 1186  
 /note="sigma-B dependent 5' end"  
 1198..1205  
 1211..1642  
 /note="ORF1"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AAB40044.1"  
 /db\_xref="PID:g556015"  
 /translation="MTIIILYSVALIAVFLVLYLSTLKSLQTLTKNVASTLEGL  
 EGMKITTETAEHLKTRRLAEDIDSEKRLNTVYHVOGASVQDNTSMKQAC  
 SVSASVREMODTINOVGRSTWSQAMEIWEKOKKKSAL"  
 1660..1665  
 1670..2128  
 /note="similar to plant water stress proteins: ORF2"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AAB40045.1"  
 /db\_xref="PID:g556016"  
 /translation="MMSKDGINSKDFLIGTIGIIGATTALFLAPKSGKELRDIDS  
 GAVALEDKRDKMDAKKGTGYSTAKKSTNTQIVADQSGQITNNKYKDLRDSKS  
 DKTDSSTANQDREHMQADETKQVLOTKEVDKDLKQAKQELNR"  
 2135..2139  
 2152..2320  
 /note="open reading frame extends an additional 53 codons  
 (including the TGA stop) into the overlapping sequence  
 reported by Bojotin et"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AAB40046.1"

CDS

prim\_transcript.1186  
 /note="sigma-B dependent 5' end"  
 1198..1205  
 1211..1642  
 /note="ORF1"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AAB40044.1"  
 /db\_xref="PID:g556015"  
 /translation="MTIIILYSVALIAVFLVLYLSTLKSLQTLTKNVASTLEGL  
 EGMKITTETAEHLKTRRLAEDIDSEKRLNTVYHVOGASVQDNTSMKQAC  
 SVSASVREMODTINOVGRSTWSQAMEIWEKOKKKSAL"  
 1660..1665  
 1670..2128  
 /note="similar to plant water stress proteins: ORF2"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AAB40045.1"  
 /db\_xref="PID:g556016"  
 /translation="MMSKDGINSKDFLIGTIGIIGATTALFLAPKSGKELRDIDS  
 GAVALEDKRDKMDAKKGTGYSTAKKSTNTQIVADQSGQITNNKYKDLRDSKS  
 DKTDSSTANQDREHMQADETKQVLOTKEVDKDLKQAKQELNR"  
 2135..2139  
 2152..2320  
 /note="open reading frame extends an additional 53 codons  
 (including the TGA stop) into the overlapping sequence  
 reported by Bojotin et"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AAB40046.1"

CDS

RBS

/codon\_start=1  
 /transl\_table=11  
 /protein\_id="AAB40044.1"  
 /db\_xref="PID:g556015"  
 /translation="MTIIILYSVALIAVFLVLYLSTLKSLQTLTKNVASTLEGL  
 EGMKITTETAEHLKTRRLAEDIDSEKRLNTVYHVOGASVQDNTSMKQAC  
 SVSASVREMODTINOVGRSTWSQAMEIWEKOKKKSAL"  
 1660..1665  
 1670..2128  
 /note="similar to plant water stress proteins: ORF2"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AAB40045.1"  
 /db\_xref="PID:g556016"  
 /translation="MMSKDGINSKDFLIGTIGIIGATTALFLAPKSGKELRDIDS  
 GAVALEDKRDKMDAKKGTGYSTAKKSTNTQIVADQSGQITNNKYKDLRDSKS  
 DKTDSSTANQDREHMQADETKQVLOTKEVDKDLKQAKQELNR"  
 2135..2139  
 2152..2320  
 /note="open reading frame extends an additional 53 codons  
 (including the TGA stop) into the overlapping sequence  
 reported by Bojotin et"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AAB40046.1"



/translation "MKKLIIFAEFGILDSIGLSAGAYVTSLSLLAIGITLPAPVA  
 ASHLEAVYVTAASGVSHKEGNADKDTOLVLPGSIGALGAAPLSOLDGAPAKPA  
 ISFLILIGLGVYIRLFEOYKPALEKRVHPLNROSIPLGVIAEAPATGGGGGPPV  
 TPLLRKGISSPKVYVDTVSPFAIVASAGCLISLGEVDVMIWVFSMAGGIIIT  
 APFAAMLVQCFHFQHLQAGVLYGCFIILVNATLIINEMIANVAPHELIITAI  
 LMSANL  
 LEVYSIGNNNIATVADYLER  
 complement(1020. .2270)  
 /gene="hipo"  
 complement(1020. .2270)  
 /gene="hpo"  
 complement(1020. .2270)  
 /EC\_number="3.5.1.32"  
 /codon\_start=-1  
 /transl\_table=-1  
 /product="hipurate hydrolase"  
 /protein\_id="CA814907.1"  
 /db\_xref="PIR:e1185802"  
 /db\_xref="PIR:g2635413"  
 /db\_xref="GI:2635413"  
 /db\_xref="SPREMBL:O34980"  
 /translation "MSIDYVRNNEGSPYQOTGNDILTLKEESNPVNLSTLEKOLIG







Db	50062	FASTGCCCTT	50073
CP	122	GCTACCATATTGTGTGAGACAGGAAGTGATCTATAAACCACCATCCACATTCACATCA	63
Db	50062	AAAGTGTCCCTT	50073
CP	62	CAGCAGTACTCT	51
RESULT	4		
LOCUS	AF008220	220060 bp	DNA
DEFINITION	Bacillus subtilis rrmB-dnaB genomic region.		BC1
ACCESSION	AF008220		
VERSION	92293135		
KEYWORDS	AF008220.1 GI:2293135		
SOURCE			
ORGANISM	Bacillus subtilis.		
REFERENCE	Bacillus subtilis.		
AUTHORS	Eubacteria; Firmicutes; Low G-C gram-positive bacteria;		
TITLE	Bacillaceae; Bacillus.		
REFERENCE	1 (bases 1 to 7430)		
AUTHORS	Green,C.J., Stewart,G.C., Hollis,M.A., Vold,B.S. and Bott,K.F.		
TITLE	Nucleotide sequence of the Bacillus subtilis ribosomal RNA operon, rrmB		
JOURNAL	Gene 37 (1-3), 261-266 (1985)		
MEDLINE	86031361		
REFERENCE	2 (bases 153210 to 153762)		
AUTHORS	Conners,M.J., Mason,J.M. and Setlow,P.		
TITLE	Cloning and nucleotide sequencing of genes for three small, acid-soluble proteins from Bacillus subtilis spores		
JOURNAL	J. Bacteriol. 166 (2), 417-425 (1986)		
MEDLINE	86195826		
REFERENCE	3 (bases 213161 to 218473)		
AUTHORS	Ogasawara,N., Moriya,S., Mazza,P.G. and Yoshikawa,H.		
TITLE	Nucleotide sequence and organization of dnaB gene and neighbouring genes on the Bacillus subtilis chromosome		
JOURNAL	Nucleic Acids Res. 14 (24), 9989-9999 (1986)		
MEDLINE	87117549		
REFERENCE	4 (bases 200404 to 201481)		
AUTHORS	Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.		
TITLE	Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis		
JOURNAL	J. Bacteriol. 169 (7), 2913-2916 (1987)		
MEDLINE	87250247		
REFERENCE	5 (bases 201234 to 203212)		
AUTHORS	Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.		
TITLE	Nucleotide sequence of the Bacillus subtilis phoR gene		
JOURNAL	J. Bacteriol. 170 (12), 5935-5938 (1988)		
MEDLINE	89053932		
REFERENCE	6 (bases 142232 to 144147)		
AUTHORS	Grundy,F.J. and Henkin,T.M.		
TITLE	Cloning and analysis of the Bacillus subtilis rpsD gene, encoding ribosomal protein S4		
JOURNAL	J. Bacteriol. 172 (11), 6372-6379 (1990)		
MEDLINE	91035248		
REFERENCE	7 (bases 133624 to 134990)		
AUTHORS	Henkin,T.M., Grundy,F.J., Nicholson,W.L. and Chambliss,G.H.		
TITLE	Catabolite repression of alpha-amylase gene expression in Bacillus subtilis involves a trans-acting gene product homologous to the Escherichia coli lacI and galP repressors		
JOURNAL	Mol. Microbiol. 5 (3), 575-584 (1991)		
MEDLINE	91260441		
REFERENCE	8 (bases 140810 to 142610)		
AUTHORS	Henkin,T.M., Glass,B.L. and Grundy,F.J.		
TITLE	Analysis of the Bacillus subtilis tyrS gene: conservation of a regulatory sequence in multiple tRNA synthetase genes		
JOURNAL	J. Bacteriol. 174 (4), 1299-1306 (1992)		
MEDLINE	92138624		
REFERENCE	9 (bases 217570 to 220060)		
AUTHORS	Putzer,H., Gendron,N. and Grundberg-Menago,M.		
TITLE	Co-ordinate expression of the two threonyl-tRNA synthetase genes in Bacillus subtilis: control by transcriptional antitermination involving a conserved regulatory sequence		

JOURNAL MEDLINE REFERENCE AUTHORS TITLE	EMBO. J. 11 (8), 3117-3127 (1992)
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	9234/7349 10 (bases 134990 to 141290) Grundy, F.J., Waters, D.A., Yakova, T.Y. and Henkin, T.M. Identification of genes involved in utilization of acetate and aceto in <i>Bacillus subtilis</i> Mol. Microbiol. 10 (2), 259-271 (1993)
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	95020526 11 (bases 162129 to 164080) Grundy, F.J., Waters, D.A., Allen, S.H. and Henkin, T.M. Regulation of the <i>Bacillus subtilis</i> acetate kinase gene by CcpA J. Bacteriol. 175 (22), 7348-7355 (1993)
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	94042910 12 (bases 7009 to 15526) Kiel, J.A., Boels, J.M., Belman, G. and Venema, G. Glycogen in <i>Bacillus subtilis</i> : molecular characterization of an operon encoding enzymes involved in glycogen biosynthesis and degradation Mol. Microbiol. 11 (1), 203-218 (1994)
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	94195107 13 (bases 196487 to 200620) Jin, S. and Sonenshein, A.L. Identification of two distinct <i>Bacillus subtilis</i> citrate synthase genes J. Bacteriol. 176 (15), 4669-4679 (1994)
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	94321340 14 (bases 16985 to 19588) Abe, A., Koide, H., Kohno, T. and Watabe, K. <i>A. Bacillus subtilis</i> spore coat polypeptide gene, <i>cots</i> Microbiol. 141 (Pt 6), 1433-1442 (1995)
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	95400496 15 (bases 131934 to 133970) Bologin, A., Khazak, V., Stoyanova, N., Ratmanova, K., Yomantas, Y. and Kozlov, Y. Identical amino acid sequence of the <i>aroA(G)</i> gene products of <i>Bacillus subtilis</i> 168 and <i>B. subtilis</i> Marburg strain Microbiology 141 (Pt 9), 2219-2222 (1995)
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	96118703 16 (bases 25258 to 31212) Rowland, B., Hill, K., Miller, P., Driscoll, J. and Taber, H. Structural organization of a <i>Bacillus subtilis</i> operon encoding menaquinone biosynthetic enzymes Gene 167 (1-2), 105-109 (1995)
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	96144257 17 (bases 196487 to 200620) Jin, S., De Jesus-Berrios, M. and Sonenshein, A.L. <i>A. Bacillus subtilis</i> malate dehydrogenase gene J. Bacteriol. 178 (2), 560-563 (1996)
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	96134995 18 (bases 129888 to 132207) Varon, D., Brody, M.S. and Price, C.W. <i>Bacillus subtilis</i> operon under the dual control of the general stress transcription factor $\sigma$ B and the sporulation transcription factor $\sigma$ H Mol. Microbiol. 20 (2), 339-350 (1996)
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	96310371 19 (bases 81540 to 91690) Bower, S., Perkins, J.B., Yocum, R.R., Howitt, C.L., Rahaim, P. and Pero, J. Cloning, sequencing, and characterization of the <i>Bacillus subtilis</i> biotin biosynthetic operon J. Bacteriol. 178 (14), 4122-4130 (1996)
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	96312354 20 (bases 49093 to 51682) Yocum, R.R., Perkins, J.B., Howitt, C.L. and Pero, J. Cloning and characterization of the <i>metE</i> gene encoding S-adenosylmethionine synthetase from <i>Bacillus subtilis</i> J. Bacteriol. 178 (15), 4604-4610 (1996)
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	96345628 21 (bases 100760 to 102298) Kappes, R.M., Kempf, B. and Bremer, F. Three transport systems for the osmoprotectant glycine betaine operate in <i>Bacillus subtilis</i> : the osmoprotectantization of <i>Opd</i>







	Intron	2427..6475 /gene="Cpl" 4346..4353 /gene="Cpl"
	misc.feature	/note="insertion site of p(Caspar)(500)" /citation=[1] 6476..6690 /gene="Cpl"
	exon	/number=3 6691..6750 /gene="Cpl"
	Intron	6751..7707 /gene="Cpl"
	exon	/number=4 join(<8110..9300,9370..>9532) /product="phenylalanyl tRNA synthetase"
mRNA		join(8110..9300,9370..>9532) /note("potential orf") /codon_start=1 /product="phenylalanyl tRNA synthetase"
CDS		/protein_id="AAB65750.1" /db_xref="PID:g2305222" /db_xref="GI:2305222"
		/translation="MULTIRVOCARHMLKSNRCLASSAPAKSPSSPPROLETVSQSTV TDCRTNTPTKILSVGNKKHLIOTDHP.SITFORLVNFTYGAQRNRORPLESYDNLD PVTVQONEDNLIPADHVSRQSDCYIIINOHLRHHTAHNOVELISGLDENLWVO EYNRRDEIDSTHYFVFQADAVRLVTDKLEERNPGDELFEETSGTLADKLILPH SWTKPNSPATRYAFKLMHEMKHVLGLTDLDFGPRIKYRWVDYTFPTOPSEMLE YFNDOENEGKLNLRHEILORSOSHOISIGAFVGELERLAWLFDDILPEMNDSD GFLSOSEKDLNLPKPKPIISHYPOCTINDLSFWLPODEVAGSPPNFOLVRSVAGA DMHQISLVDKFKHKPTGKSSVCRIYRHRMERLTLOAEVNEIHKOIASASYSDFNVG IX"
BASE COUNT	2929 a 2357 c 2282 g 3046 t	158 others
ORIGIN		
Query Match	6.4%; Score 42;	DB 21; Length 10772;
Best Local Similarity	20.0%;	Pred. No. 3.25e+08;
Matches	31; Conservative	70; Mismatches 53; Indels 1; Gaps 1;
Db	1691 AKRMWYMAWTMTMMKMMWTTWKMMAMKTIRYWMWMMKMWTSPRTTSAMMMWYTSTWTK	1750
OY	385 ATTTTAATGAATTTTCAGAAAGTTTAAGTAAACAGACTGTGATTCTTAGTGGAAT	444
Db	1751 YWMAVAMKMWMTTRTWANMANASMARWMTKTSAAAYASAMRKMRWMAVRAKKTMWMA	1810
OY	445 TTT-TGCATCAATTAGAAAATACTGGCGCATTAACGATACAGATTAAATTGATAANA	503
Db	1811 WKMRKKAAMWTRWMTMYMTTAAAMPACGTTCTATT	1845
OY	504 TTGAAGTGCAATCGTTAATTATGAAGATTCTATT	538
LOCUS	7	
FEATURE	AF012089	10772 bp DNA INV 05-AUG-1997
DEFINITION	Drosophila melanogaster cysteine proteinase-1 (Cpl) gene, complete cds, and phenylalanyl tRNA synthetase gene, partial cds.	
ACCESSION	AF012089	
NID	G2305220	
VERSION	AF012089.1	GI:2305220
KEYWORDS		
SOURCE	fruit fly,	
ORGANISM	Drosophila melanogaster	
REFERENCE	Eukaryotes: mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata: Hexapoda; Insecta: Pterygota; Diptera: Brachycera; Muscomorpha: Ephydroidea; Drosophilidae; Drosophila. 1 (bases 4546 to 4553)	
AUTHORS	Gray,Y.H., Tanaka,M.M. and Sved,J.A.	



TITLE P-element-induced recombination in *Drosophila melanogaster*: hybrid  
JOURNAL element insertion  
MEDLINE Genetics 144 (4), 1601-1610 (1996)  
97132596  
REFERENCE 2 (bases 1 to 10772)  
AUTHORS Gray, J.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.  
TITLE Structure of the cysteine proteinase (Cp1) gene of *Drosophila melanogaster* and associated mutational effects  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 10772)  
AUTHORS Gray, J.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.  
TITLE Direct Submission  
JOURNAL Submitted (30-JUN-1997) School of Biological Sciences, University of Sydney, Biology A12, Sydney University, NSW 2006, Australia  
FEATURES  
SOURCE 1. 10772  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
join(872..1000,2310..2426,6476..6690,6751..7707)  
/gene="Cp1"  
/product="cysteine protease"  
872..7707  
/gene="Cp1"  
872..1000  
/gene="Cp1"  
/number=1  
1001..2309  
/gene="Cp1"  
2310..2426  
/gene="Cp1"  
/number=2  
join(2328..2426,6476..6690,6751..7462)  
/gene="Cp1"  
/codon\_start=1  
/product="cysteine proteinase-1"  
/protein\_id="AAB5749.1"  
/db\_xref="PID:92305221"  
/db\_xref="GI:2305221"  
/translation="MRTAVLLPLALLAQAQVAFADVMEHTEFLHKKYQDET  
EERFKTFNNKHKIAKHNORFAGKVSFELANAKYADLHFERFLMNGPNTLHK  
QLRADESKGYTFISPAHYTLPSVDMRTKGAATYAKDGHCSGASSTGLBEO  
HFKRSGLVLSLEQNLVDCSTKYGNNCGMLDANARYIAIDSHESFQYSGVYN  
DSCHFNKGTGATDRGFTDIPQDEKMAEVAIVGVSVAIDASHESFQYSGVYN  
EPQCDQONLDHGVLYVGFGTDESEGYWLVNSWGTWMDKGFIKMLRNKNGCGIAS  
ASSYPLV"  
2427..6475  
/gene="Cp1"  
4346..4533  
/gene="Cp1"  
/note="insertion site of p(Caspar)(50C)"  
/citation=[1]  
6476..6690  
/gene="Cp1"  
/number=3  
6691..6750  
/gene="Cp1"  
6751..7707  
/gene="Cp1"  
/number=4  
join(<8110..9300,9370..>9532)  
/product="phenylalanyl tRNA synthetase"  
join(8110..9300,9370..>9532)  
/note="potential orf"  
/codon\_start=1  
/product="phenylalanyl tRNA synthetase"  
/protein\_id="AAB5750.1"  
/db\_xref="PID:92305222"  
/db\_xref="GI:2305222"  
/translation="MLTLRVQGARHMLKSTRCLASAAAPKSPSPQLEVGSTYA  
TDGNTNTPKLISVGNKHLQTDHPISIRQIVNTFYATIRKQKNDPLFSEYDQAN  
PVYVQGNFNDLLIPADHVSROKSDCYINQOHLRAHTAHQVELISGDLNPLVVG  
EYARDEIDSHLIPADHVSROKSDCYINQOHLRAHTAHQVELISGDLNPLVVG  
SSWTKPNSPATIRRAVKIMEHMKHVLGLTKDLFGPRIKYRWDTYFPFPQSPWLEI

YFKDNNLEVLGCGIMRHEILQSRGSHQSIGYAGVGLERLAWLFDIPDIRLEWNSD  
GFLSPSEKDLNLPKPKYKPSHYPQCTNDLSEFWLPQDIEVDAGFSPNDEYDLRSVAG  
IX  
DMEQJSLVDKFKHPRKTKSSVCFRIYVHMETTLQAEVNEHROIAASVDSFVNO

BASE COUNT 2929 a 2357 c 2282 g 3046 t 158 others  
ORIGIN

Query Match 5.3% Score 35; DB 21; Length 10772;  
Best Local Similarity 19.0%; Pred. No. 1.88e-04;  
Matches 23; Conservative 60; Mismatches 36; Indels 2; Gaps 2;

Db 1706 MWMTTWKMAKMYRTWMMKMYTSRTTSSAMWYTWSTWTKY-WMAVAMKMMWTR 1764  
Cp 651 TATTAACAGCATTTTCATGCCCTAATTATCTAAATATGCATTTTGAATA 592

Db 1765 TWAMMAASMAKMKWKSAAAYASAMKMKWMAVYRAMK-TMMMAVKKMKRAMTWR 1823  
Cp 591 TCACCTGCACCCATTAATAACACAGCATTTATCAATTTCTATATACATTAATGAA 532

Db 1824 W 1824  
Cp 531 T 531

RESULT 8  
LOCUS MYU87256 1056 bp DNA MM 02-JAN-1999  
DEFINITION Mustela vison GT dinucleotide repeat, chromosome 1q.  
ACCESSION U87256  
NID 94099442  
VERSION U87256.1 GI:4099442  
KEYWORDS  
SOURCE American mink.  
ORGANISM  
Mustela vison  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Carnivora; Fissipedia; Mustelidae; Mustela.  
REFERENCE 1 (bases 1 to 1056)  
AUTHORS Brusgaard, K., Shukri, N.M., Malchenko, S., Koroleva, I. and Loh, O.  
TITLE Direct Submission  
JOURNAL Submitted (27-JAN-1997) Breeding and Genetics, Danish Institute of  
Animal Science, Blichersalle K25, Tjele 8830, DK  
FEATURES  
source 1..1056  
/organism="Mustela vison"  
/db\_xref="taxon:9667"  
/map=1q  
/note="primers: 1167F: agccctgcatctactctt, 1167R:  
gagatcttaccgctgtgag"  
98..119  
/standard\_name="1167F"  
primer\_bind complement(300..320)  
primer\_bind /standard\_name="1167R"  
BASE COUNT 211 a 221 c 210 g 225 t 189 others  
ORIGIN

Query Match 4.7% Score 31; DB 23; Length 1056;  
Best Local Similarity 34.5%; Pred. No. 2.01e-02;  
Matches 30; Conservative 29; Mismatches 28; Indels 0; Gaps 0;

Db 706 YARRKMYGTGAAMASATAYSTGAGTTBMYTSKMBKYYKTKGMHNCAMKYGKGGCCY 765  
Cp 401 CAATATGATTTAAAGATGCTGTGTTCTAAGAAAGAGTGTTGTTGAATACGCAACA 342

Db 766 CWCHTTATGBGATNMGTTMMWSTG 792  
Cp 341 CTCTTATGATGATATTTCTTCGTG 315

RESULT 9  
LOCUS AE001180 15079 bp DNA BCT 15-DEC-1997  
DEFINITION *Borrelia burgdorferi* (section 66 of 70) of the complete genome.  
ACCESSION AE001180 AE000783  
NID 92688755



KEYWORD	AE001180.1	GI:2688755	
SOURCE	Lyme disease spirochete.		
ORGANISM	Borrelia burgdorferi		
REFERENCE	Eubacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group.		
AUTHORS	1 (bases 1 to 15079) Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J.D., Weidman, J., Uterback, T., Wathey, L., McDonald, L., Attiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C.		
TITLE	Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi		
JOURNAL	Nature 390 (6660), 580-586 (1997)		
MEDLINE	98065943		
REFERENCE	2 (bases 1 to 15079) Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J.D., Weidman, J., Uterback, T., Wathey, L., McDonald, L., Attiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C.		
AUTHORS	Direct Submission		
TITLE	Submitted (12-DEC-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA		
JOURNAL	Location/Qualifiers		
FEATURES	1. 15079		
SOURCE	/organism="Borrelia burgdorferi"		
gene	/db_xref="taxon:139"		
CDS	68..244		
gene	/gene="Bb0813"		
CDS	68..244		
gene	/note="hypothetical protein; identified by Glimmer; putative"		
CDS	/codon_start=1		
gene	/transl_table=11		
CDS	/product="B. burgdorferi predicted coding region Bb0813"		
gene	/protein_id="AAC6716.1"		
CDS	/db_xref="PID:g268871"		
gene	/db_xref="GI:2688771"		
CDS	/translation="MRNGILYAILLYVFVFLMFCFAVIFDITITINPLMFLSLGL LPSILIFWVFL"		
gene	322..1656		
CDS	322..1656		
gene	/gene="Bb0814"		
CDS	/note="similar to PID:606198 GB:U00096 PID:1789656 percent identity: 33.64; identified by sequence similarity; putative"		
gene	/codon_start=1		
CDS	/transl_table=11		
gene	/product="pantothenate permease (panF)"		
CDS	/protein_id="AAC67168.1"		
gene	/db_xref="PID:g2688763"		
CDS	/db_xref="GI:2688763"		
gene	/translation="MLNKYFLANRNINFINVALLFPSSYISASSFISGPAVYKGL SFLLATIOIPTLIYFIIVGRLNRESKRNAINIDYIRHRESDFALMSGVLI FEGMFLISAGLIGAKLIEVGMGIDVYGLTAPALPIYVFFGSKFAYATDLOGF LMYSSYILRSKMLDGGGINNFKTATSLDLSKSLPSNADLKOYITISWILIGIG ILDQPOLINNFIAFKODNAIKFSLPSTIISLIYLMHLIGFPAIILFDPDISPDYK VLVNVALVMPFSCFMFETGLISAIMSTVSDNLTITVLSISIFIKEDLKEDYKIG RIMISNIFPILILILIFSLFPNFIIVFAFAGALEVSEFPPIIVGGLYLVNSKIA FASVFLGLIYFLISILFGLNIMFWHPVPSFISITFLVAVFPCKNSKVC"		
gene	1666..2559		
CDS	1666..2559		
gene	/gene="Bb0815"		
CDS	1666..2559		











LOCUS HUAC004787 216021 bp DNA PRI 24-JUL-1998  
DEFINITION Homo sapiens Chromosome 16 BAC clone C19875K-A-952F10, complete  
sequence.  
ACCESSION AC004787  
NID 93337381  
VERSION AC004787.1 GI:3337381  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 216021)  
ADAMS,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,  
Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.  
Homo sapiens Chromosome 16 BAC clone C19875K-A-952F10  
Unpublished  
2 (bases 1 to 216021)  
ADAMS,M.D. and Loftus,B.J.  
Direct Submission  
Submitted (02-JUN-1998) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA, Email:  
bjloftus@tigr.org  
3 (bases 1 to 216021)  
ADAMS,M.D. and Loftus,B.J.  
Direct Submission  
Submitted (24-JUL-1998) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
On Jul 24, 1998 this sequence version replaced gi:3241936.  
Address all correspondence to: Mark Adams The Institute for Genomic  
Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail  
address: humgen@tigr.org. The orientation of the sequence is from  
SP6 end to 3' end. Genes were identified by a combination of five  
methods including: XGRail (available by anonymous ftp from  
arthur.epm.ornl.gov), GeneFinder (Phil Green, University of  
Washington), Genscan (Chris Burge,  
http://genomic.stanford.edu/~chris/GENSCAN.html) searches of the  
complete sequence against a peptide database, and the Human gene  
index database at tigr (http://www.tigr.org/tdb/hgi.html).  
Genes without peptide homology having spliced EST hits are termed  
'Unknown gene product'. Genes encoding tRNAs are predicted by  
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).  
Location/Qualifiers  
1. 216021  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16q21-22"  
/clone="A-952F10"  
2765..27872  
/note="7766, STS1-CSRL-27g3-uA/CSRL-27g3-uZ, Chr. -, Homo  
sapiens"  
/db\_xref="dbSTS:G02280"  
73826..73943  
/note="7608, STS1-CSRL-24g1-uA/CSRL-24g1-uZ, Chr. -, Homo  
sapiens"  
/db\_xref="dbSTS:G02122"  
175801..175945  
/note="16084, CHLC-GCT10B02, Chr. -, Homo sapiens"  
/db\_xref="dbSTS:G09703"  
175810..175945  
/note="16316, CHLC-GCT15C04, Chr. -, Homo sapiens"  
/db\_xref="dbSTS:G09935"  
199463..199572  
/note="9824, WI-3555, Chr. 16, Homo sapiens"  
/db\_xref="dbSTS:G04338"  
BASE COUNT 60960 a 51778 c 49172 g 53987 t 124 others  
ORIGIN  
Query Match 4.7% Score 31; DB 31; Length 216021;  
Best Local Similarity 13.1%; Pred. No. 2 01e-02;  
Matches 14; Conservative 55; Mismatches 38; Indels 0; Gaps 0;  
Db 1451 AARRGKKRYMKSMBRRRRGAGKRRKRYYYCYCYCYCYCYCGRAMMAAAYTKRRS 1510

LOCUS 12 U32794 12085 bp DNA BCT 09-SEP-1998  
DEFINITION Haemophilus influenzae Rd section 109 of 163 of the complete  
genome.  
ACCESSION U32794.L42023  
NID g1574694  
VERSION U32794.1 GI:1574694  
KEYWORDS  
SOURCE  
ORGANISM Haemophilus influenzae Rd.  
Haemophilus influenzae Rd.  
Eubacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus.  
REFERENCE  
AUTHORS 1 (bases 1 to 12085)  
Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A.,  
Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A.,  
Merrick,D.M., McKenney,K., Sutton,G.G., FitzHugh,W., Fields,C.A.,  
Gocayne,J.D., Scott,J.D., Shirley,J., Liu,L.I., Glodek,A.,  
Keller,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E.,  
Cotton,M.D., Uterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M.,  
Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L., Fraser,C.M.,  
Geoghegan,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M.,  
Smith,H.O. and Venter,J.C.  
Whole-genome random sequencing and assembly of Haemophilus  
influenzae Rd  
Science 269 (5223), 496-512 (1995)  
9530630  
2 (bases 1 to 12085)  
Tatusov,R.L., Mshnegian,A.R., Bork,P., Brown,N.P., Hayes,W.S.,  
Borodovsky,M., Rudd,K.E. and Koonin,E.V.  
Metabolism and evolution of Haemophilus influenzae deduced from a  
whole-genome comparison with Escherichia coli  
Curr. Biol. 6 (3), 279-291 (1996)  
96398784  
3 (bases 1 to 12085)  
White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.  
Direct Submission  
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
4 (bases 1 to 12085)  
White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.  
Direct Submission  
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
The H. influenzae sequence has been updated by R. Fleischmann. New  
database matches have been assigned, product names have been  
improved, and a number of frame shifts have been corrected. We  
gratefully acknowledge the work of Tatusov et. al. We have  
incorporated their annotation into the /notes fields of the  
corresponding H. influenzae genes  
5 (bases 1 to 12085)  
White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,  
Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.  
Direct Submission  
Submitted (28-MAY-1998) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
The whole genome was shifted by 588 nucleotides for a new start  
on Oct 1, 1996 this sequence version replaced gi:121895.  
Location/Qualifiers  
1. 12085  
/organism="Haemophilus influenzae Rd"  
/db\_xref="taxon:71421"  
99..1526  
/gene="H1139"  
99..1526  
/gene="H1139"  
FEATURES  
source  
gene  
CDS



```

/note="similar to GB:D10483 SP:P17952 PID:216505 PID:40859
PID:42056 percent identity: 68.16; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylmuramate--alanine ligase (murC)"
/protein_id="AAC22794.1"
/db_xref="PID:91574695"
/db_xref="GI:1574695"
/translation="MKHSHEEIRKIIPEMRVQOIHFIIGAGSGIAEILLNEGYO
ISGSDIADGVVTOFLAOGAKIYIGHAEHIEGASVVVSSAIKDDNPELVTSKRI
PIYQAOIMAEIMRRRHGIAVGTGKTTTAMISITQALDPTFNGGKYKACK
NAHGAARLLEADESDASTLHLOPMVSVVIMEDDHDITESDPEKKAITYVFL
NIPFYGLAVMCADDPVLMELVPKRGVYITGFSQADRIDEDYQGTQGYATVCP
NNERINVLVDYGGHHHTNATLAVALAKEEGJANEAILEALADFOGAGRFDLGEFI
RNGKRVLYVDYGGHHHTNATLAVALAKEEGJANEAILEALADFOGAGRFDLGEFI
SOVDALIMDYVAAAGEAPYVGSGLCSINLCKVPIVLSQSDGVLDQIIDD
GDILAOAGSVYSKISRLAESMKN"
1598..2518
/gene="H11140"
1598..2518
/note="similar to GB:D10483 SP:P07862 GB:K02668 PID:145724
PID:146030 percent identity: 59.93; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="D-alanine--D-alanine ligase (ddlB)"
/protein_id="AAC22795.1"
/db_xref="PID:91574696"
/db_xref="GI:1574696"
/translation="MNLKQEKJAVLVGTSAREVSLNSGKAVLEALLKOGYNAPID
PREYVANLKRDGFNRANFIHGRGEGDTMOGLFOJGLPYTGCGVNASALTMMKMR
TKMLKAGLPVADMKVYVTRFSELDPAVYAKIGLPMVKPSLEGGSVGLTXKAY
BELKSAVERALKEDNTILIEMLACDELTPVLDNOVLPALITPEGEGYDIEAKTIS
DNTQYFCPAGLTPERBOALSTLVKRAYDVGGRGSRIDVCMKADANGNEFLVEVNTNG
MISHSLFPKSATVGISEFQVYVLTLESL"
2518..3282
/gene="H11141"
2518..3282
/note="similar to GB:D10483 SP:P06136 GB:K02668 GB:X02821
PID:145725 percent identity: 40.62; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="cell division protein (ftsQ)"
/protein_id="AAC22796.1"
/db_xref="PID:91574697"
/db_xref="GI:1574697"
/translation="MNLKRRTPONIRREGOKPKYFPHIRAFVALLGVFELLGVENW
OSILEKMDKPIASAFALGONMFTADDIKESILKMGELKSGODVAPIDQIADLP
WYKGAIVKRMKPNRSLIYSEKQPAFNMQNDPVLIDGIVLPSPVRLAKVLPYLG
PYQSLKVLETWNOIYINLKSNNIMAKGINIDRGAMVOVDNDVLKLRGDMKSKL
EEFVIVYQIDVDPENKKIDYIDRLRYTAGAAGVMDR"
3301..4578
/gene="H11142"
3301..4578
/note="similar to GB:D10483 SP:P06137 GB:K02668 GB:M36531
GB:X02821 percent identity: 52.84; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="cell division protein (ftsA)"
/protein_id="AAC22797.1"
/db_xref="PID:91574698"
/db_xref="GI:1574698"
/translation="WYKGEETVIGLEVGTGSKVVVVGEVEPDGVVNVLYGSGPSK
GIDRGSITLDAVVGSIORAIABESMACCOISMVTLATGTHGSIOLNSGSGVLAES
EYTOEIDSLHTASIKRPEGLSLHVIPOQYAYNDROMKNIPNIGLQGVRLKAQVHL
IACHODMNNLKAVERGLOVDKVFESFAFTHSVLEDERDGLVCLIDFGAGTMNV
MYTNGALRFSKVIIPYAGNIVTNDIAHACTISRAAERIKVYASAFYPARLHGDKKI

```

```

EVAISIGRAPSLSIKSDISLTSARVTELLGVKDELDKKAELAEHRIKELIAGVY
ITGGGAQIEDLEKCAASNFRQCVRIASPLNTGLDVNRPQYSTVGLLOLYNNSND
DDLISGSDSEGTFEESIMQGVKRIKVRKSEF"
4662..5927
/gene="H11143"
4662..5927
/note="similar to GB:D10483 SP:P06138 GB:X02821 PID:216509
PID:40863 percent identity: 67.53; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="cell division protein (ftsZ)"
/protein_id="AAC22798.1"
/db_xref="PID:91574699"
/db_xref="GI:1574699"
/translation="WLYREYEPYDNFNESGALIVYGVGGGGAHVNMNMNKKQEM
GCTVGESSLTSEHGRIVYAVNDDAOLKRSOYQOYVIGGETTKELGAGANPNIG
RKAADDQDEIRKMLEGADWVFIAAGMGGGTGAAPVVAIRAEKILITVAAYVTKP
TFEGKRRMOFAELGIKIDLSQYVDSWIIIPNOQIQVLPKNAKLIDAFMAANDVLRNSY
MGISDMITSPGLINVDADVATVMSVQAGMIGFSAVEGAGAAEAAVLAAYNDL
LEKIDLSAAGIIVNITGADNLVEEFENIIGETIGSFASEEATVYVGTSLVPMSEDI
RVTIYATGIGSIAGNERIQVYRQGLSTONIGEGRVNIVPELHRESVEVSRITASEY
QRPDLKPTIDRLAEFRKNFNPQREEN"
5966..6883
/gene="H11144"
5966..6883
/note="similar to GB:D10483 SP:P07652 PID:145848
PID:216510 PID:40864 percent identity: 77.30; identified
by sequence similarity: putative"
/codon_start=1
/transl_table=11
/product="DDP-3-O-(3'-hydroxy-myristoyl) N-acetylglucosamine
deacetylase (lpxC)"
/protein_id="AAC22799.1"
/db_xref="PID:91574700"
/db_xref="GI:1574700"
/translation="MIKORTLKOSIKVYGVGLHSGEKVTLRPAMPNTGVVYRTDL
NPAYFAPADPSVSDTMCCTALINDGCRVTSVHNAALAGIDNIIIEVDAPETP
IMDGSASFYILLDAGIEONAKKFRIRIOYRVAVEDGDKMAEKRPYNGRLDPTID
FDHPAIGADVNRNENFSAQAFVQDISARTEGFKKDEYVLSQGLVGGSLDNAIVL
DDYRLINDEGDFRDELVRHKLDAIGDLYMAGYNIIGDFKAYKSGHLLNKLRLAVL
ANOEAWEFEVFEEDKAOVPOGYAPVQVLI"
7010..8167
/gene="H11145"
7010..8167
/note="similar to SP:Q02286 GB:M74134 GB:X60420 PID:148468
PID:43344 percent identity: 55.32; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="Chorismate mutase / prephenate dehydratase
(pheA)"
/protein_id="AAC22800.1"
/db_xref="PID:91574701"
/db_xref="GI:1574701"
/translation="WALELSDIRQOITQIDRSLLKLSERHRLAFDVVRSEIKOSL
RDVEREODQLDELVOFALENENYOLEAOYTSIFOKIIDSVLTQOQVYLOKNTLRNO
NGLHIFLEKRGSGSYLAARNYAAARQYKQSEFQVEKFGADGVLPLEN
TTSGAINEVYDLQHTDLSVGLAFLPKHCVLVNDKDLNOIDITVLSHPQVIOOCSO
FIHSLDRVHIEYCESSHAMQVLVASLNKPNIAALGNEGGLTGLSVLKTINADENN

```

Note: remainder of annotations omitted.

```

Query Match      4.5%; Score 30; DB 18; Length 12085;
Best Local Similarity 60.4%; Pred. No. 6,23e-02;
Matches 87; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
Db 1125 GTTGATGATTATGTCATATCCACAGAGTAGGCGTAAACATTAAACGGCGGAGAA 1184
11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 ||||| 11 |||||
OY 263 GTAGATGATTATGACACACATCCAGAGAAATTAGTACATATGACACAGCAGGAAG 322

```



Db	1185	GGCGGGGAGTAACGATTTGATGATTTTCCACCGGATCGTCTGCGACGCGT	1244
QY	323	AAATATCCACATMAAGAAGTTGTTGCAGTATTTTCAACACACACTTCTCTAGAACACAA	382
Db	1245	GATTTGTTTGATGATTTTGTACAA	1268
QY	383	GCATTTTAAATGATTTGCAGAA	406
RESULT	13		
LOCUS	AF063866	236120 bp	DNA
DEFINITION	Melanoplus sanguinipes entomopoxvirus, complete genome.		
ACCESSION	AF063866		
NID	94049647		
VERSION	AF063866.1	GI:4049647	
KEYWORDS			
SOURCE	Melanoplus sanguinipes entomopoxvirus.		
ORGANISM	Melanoplus sanguinipes entomopoxvirus		
REFERENCE	1 (bases 1 to 236120)		
AUTHORS	Afonso,C.L., Tulman,E.R., Lu,Z., Oma,E., Kutish,G.F. and Rock,D.L.		
TITLE	The genome of Melanoplus sanguinipes entomopoxvirus		
JOURNAL	J. Virol. 73 (1), 533-552 (1999)		
MEDLINE	99102612		
REFERENCE	2 (bases 1 to 236120)		
AUTHORS	Afonso,C.L., Tulman,E.R., Lu,Z., Oma,E., Kutish,G.F. and Rock,D.L.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-MAY-1998) Microbiology, Plum Island Animal Disease Center, U.S. Dept. Agriculture, Agricultural Research Service, P.O. Box 848, Greenport, NY 11944-0848, USA		
FEATURES			
SOURCE	1. 236120		
	/organism="Melanoplus sanguinipes entomopoxvirus"		
	/isolate="Tucson"		
	/db_xref="taxon:83191"		
	1. 7201		
repeat_region	/note="left inverted terminal repeat"		
	/rpt_type="inverted"		
gene	/evidence="experimental"		
	/complement(2443..2631)		
	/gene="MSV001"		
CDS	/complement(2443..2631)		
	/gene="MSV001"		
	/codon_start=1		
	/product="ORF MSV001 62 amino acid repeat gene family protein"		
gene	/protein_id="AAC97866.1"		
	/db_xref="PID:94049906"		
	/db_xref="GI:4049906"		
	/translation="MFLYLSINQINNVSISISNTQSMFLYLSINQINNVSISISNT QSMFLYLSINQINNVSISISNT"		
	/complement(6051..6320)		
	/gene="MSV002"		
CDS	/complement(6051..6320)		
	/gene="MSV002"		
	/codon_start=1		
	/product="ORF MSV002 89 amino acid repeat gene family protein"		
gene	/protein_id="AAC97867.1"		
	/db_xref="PID:94049907"		
	/db_xref="GI:4049907"		
	/translation="MAEKRNKVKVLEPDVNSLSITTEALONEIQKSIDQSNILKQ NISEDIMNEVEKESQVETLIMLKNIYVISTIEKIDNVNLSIS"		
	/complement(6347..6814)		
	/gene="MSV004"		
CDS	/complement(6347..6814)		
	/gene="MSV004"		
	/codon_start=1		
	/product="ORF MSV004 155 amino acid repeat gene family protein"		
	/protein_id="AAC97868.1"		
	/translation="MYSCKMEIDLECVCYCKNCNVSVOYTGICINDCPINIKLDC CNCKNINFEVDVCGECYANKIVCGNDGTAKTCTITIIINDKATIEEVESSKDIFP SDFYTNINNVIISSNNKLIQIINDIKRYEKNRLLFNDHVIITQOIDEI"		
	/complement(6574..6756)		
	/gene="MSV003"		
CDS	/complement(6574..6756)		
	/gene="MSV003"		
	/codon_start=1		
	/product="ORF MSV003 60 amino acid repeat gene family protein"		
gene	/protein_id="AAC97869.1"		
	/db_xref="PID:94049909"		
	/db_xref="GI:4049909"		
	/translation="MTHLITQNVYALIVIFILLINLYIVNWKLLIFMEIYAVNVM LTKLYVAVIIQNLHVI"		
	/complement(7431..8171)		
	/gene="MSV005"		
CDS	/complement(7431..8171)		
	/gene="MSV005"		
	/codon_start=1		
	/product="ORF MSV005 hypothetical protein"		
gene	/protein_id="AAC97860.1"		
	/db_xref="PID:94049900"		
	/db_xref="GI:4049900"		
	/translation="MKMKILFKDYNKIIYKIKSLDGCYNLKKRLINDIKLSIKK LFCYNINISCVLPISIKFYITLSEIITLILLNYYIKMQINIRAKHNNKRYKQ KECLEITFQNOPEICICYKNNVFMVCHSDTIDVDELNGCKNCNINISYVYIG VLMYKRYRIKCIIGCWDVYISNIGSECKRYLFRFTKRCIOCMFIDELNWKFCG HYCYFYCYWGYWYIYPLFDYNG"		
	/complement(9825..11327)		
	/gene="MSV008"		
CDS	/complement(9825..11327)		
	/gene="MSV008"		
	/codon_start=1		
	/product="ORF MSV008 leucine rich repeat gene family protein, similar to Amsacta moorei entomopoxvirus Q3 ORF SW:P28854"		
gene	/protein_id="AAC97859.1"		
	/db_xref="PID:94048899"		
	/db_xref="GI:4049899"		
	/translation="MNNINLYCKNIIKLYLNDNDKNNIGIYELEFKYNNILKPIYV KYFTYVLYEIKNLNCSDCNITDFELESILNLEIIDSYNKNSIYCKLSLLKMLD NCNECKIIDPFKLESILNLEIIDSYNKNSIYCKLSLLKMLDNCNECKIIDPFKLES ILNLEIIDSYNKNSIYECTPLISILEICISQKIIDPFKLESILNLEIIDSYNKNSIY KSNISDCKLPVSLIKNCEFCYIYDFKLELTNNKILINISYNNKNSISDCKLPVSL IKUNGCTIYDFKLEPLINLOKLINISNSISNISKDLFSLIKRSEYCDITNKR FLERLINETLTITVYNSRNSISICKLPVSLNLPNIEYDINKIFTEKLNKSLASLS SRSRNSISICKLPVSLIYDLYSCHIKNFKLEPLINLKNLISDNGDSNISKCLPDI SLEIYCSKNIKNKNFLOPLINLOKLIDISVDFEFNILNHFVYPPIN"		
	/complement(10476..10882)		
	/gene="MSV007"		
CDS	/complement(10476..10882)		
	/gene="MSV007"		
	/codon_start=1		
	/product="ORF MSV007 hypothetical protein"		
gene	/protein_id="AAC97609.1"		
	/db_xref="PID:94049649"		
	/db_xref="GI:4049649"		
	/translation="MYIVANSKILNRSKNLKFVVISQVSDNLINDIGNLHFDIEYEL ELIFNCKLINGSKNIKSIYI"		



Db 20567 ATACTGTA CTATTATTCACAAATTATG-TATATCATTAATAATCTAAATATTGTATTATA 206255











Query Match 4.08; Score 131; DB 1; Length 493;

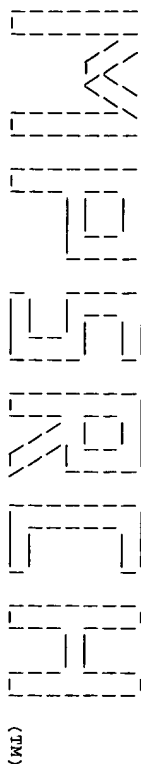






**THIS PAGE BLANK (uspto)**



\*\*\*\*\*  
  
 \*\*\*\*\* (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Merch\_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 12:35:51 1999; Maspar time 5.14 Seconds  
 381.563 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-3  
 Description: (1.660) from US09103287.seq  
 Perfect score: 660  
 N.A. Sequence: 1 AATTAAAGATTCGATGACA.....GCGTTTATATGTTTATA 660  
 Comp: TAAATTCTCAAGCCTACTCTG.....CGCAAAATATACAAATATT

Scoring table: TABLE default  
 Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 1052 seqs, 1486975 bases x 2

Post-processing: Minimum Match 08  
 Listing first 45 summaries

Database: HIV-NA8  
 1:H\_PRT 2:H\_UNA 3:H\_VIR

Statistics: Mean 11.993; Variance 3.454; scale 3.472

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	20	3.0	3440	3	HIVDJ373A Human immunodeficiency	2.69e-01
2	20	3.0	3340	3	HIVDJ373A Human immunodeficiency	2.69e-01
3	19	2.9	3033	3	HIVM2S1 Human immunodeficiency	1.15e+00
4	19	2.9	3093	3	HIVM2S1 Human immunodeficiency	1.15e+00
5	19	2.9	1459	2	HIVV1415 Human immunodeficiency	1.15e+00
6	18	2.7	9215	2	HIVMDG81 Human immunodeficiency	4.46e+00
7	18	2.7	9229	3	HIVLAI Human immunodeficiency	4.46e+00
8	18	2.7	9229	3	HIVBRU Human immunodeficiency	4.46e+00
9	18	2.7	9794	3	SIVAGM155 Human immunodeficiency	4.46e+00
10	18	2.7	9795	3	HIVTH475A Human immunodeficiency	4.46e+00
11	18	2.7	9892	3	SIVSTM Human immunodeficiency	4.46e+00
12	17	2.6	276	3	HIVU08638 Human immunodeficiency	1.55e+01
13	17	2.6	276	3	HIVU08636 Human immunodeficiency	1.55e+01
14	17	2.6	314	3	HIVU04917 Human immunodeficiency	1.55e+01
15	17	2.6	334	3	HIVM12199 Human immunodeficiency	1.55e+01
16	17	2.6	345	3	HIVU08765 Human immunodeficiency	1.55e+01
17	17	2.6	345	3	HIVU08764 Human immunodeficiency	1.55e+01
18	17	2.6	562	3	HIVTAITM1 Human immunodeficiency	1.55e+01
19	17	2.6	784	3	HIV2FA1 Human immunodeficiency	1.55e+01
20	17	2.6	1299	3	HIVRET7 Human immunodeficiency	1.55e+01

21	17	2.6	1468	3	HIVV1557 Human immunodeficiency	1.55e+01
22	17	2.6	1471	3	HIVV15251 Human immunodeficiency	1.55e+01
23	17	2.6	1474	3	HIVSE3651 Human immunodeficiency	1.55e+01
24	17	2.6	1474	3	HIVB2200 Human immunodeficiency	1.55e+01
25	17	2.6	1474	3	HIVSE365 Human immunodeficiency	1.55e+01
26	17	2.6	1474	3	HIVHP136 Human immunodeficiency	1.55e+01
27	17	2.6	2206	3	HSPULTR Human immunodeficiency	1.55e+01
28	17	2.6	2581	3	HIVK124A2 Human immunodeficiency	1.55e+01
29	17	2.6	2583	3	HIVZCAM4E Human immunodeficiency	1.55e+01
30	17	2.6	2649	3	HIVU08794 Human immunodeficiency	1.55e+01
31	17	2.6	2830	3	HIVU08449 Human immunodeficiency	1.55e+01
32	17	2.6	3457	3	HIVZ321 Human immunodeficiency	1.55e+01
33	17	2.6	8952	3	HTLV2 Human immunodeficiency	1.55e+01
34	17	2.6	9076	2	HIVAN Human immunodeficiency	1.55e+01
35	17	2.6	9168	3	HIVBSCG3C Human immunodeficiency	1.55e+01
36	17	2.6	9176	3	HIVELI Human immunodeficiency	1.55e+01
37	17	2.6	9202	3	VLVCG Human immunodeficiency	1.55e+01
38	17	2.6	9265	3	HIVHXB2R Human immunodeficiency	1.55e+01
39	17	2.6	9591	3	HIVCAM1 Human immunodeficiency	1.55e+01
40	17	2.6	9638	3	SIVSAM9 Human immunodeficiency	1.55e+01
41	17	2.6	9646	3	SIVMM142 Human immunodeficiency	1.55e+01
42	17	2.6	9706	3	HIVYU2 Human immunodeficiency	1.55e+01
43	17	2.6	9718	3	HIVHXB2 Human immunodeficiency	1.55e+01
44	17	2.6	9737	3	HIVSEF2 Human immunodeficiency	1.55e+01
45	17	2.6	9811	3	SIVCPZ Human immunodeficiency	1.55e+01

#### ALIGNMENTS

RESULT 1  
 LOCUS HIVDJ373A 3340 bp ds-DNA VRL 31-DEC-1993  
 DEFINITION Human immunodeficiency virus type 1, DJ373 from Djibouti,  
 proviral DNA encoding env, tat, vpu, rev, and nef genes.  
 ACCESSION L23065  
 KEYWORDS env gene; envelope glycoprotein; nef gene; rev gene; tat gene;  
 vpu gene.  
 SOURCE Human immunodeficiency virus type 1 (HIV-1), DJ373 from Djibouti,  
 PCR amplification of proviral DNA from cocultivated PMCs  
 ORGANISM Human immunodeficiency virus type 1  
 Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;  
 Retroviridae; Lentivirinae.  
 REFERENCE 1 (bases 1 to 3340)  
 Louwagie,J.J., Janssens,W., Mascola,J.J., Fischer,C.L., Hegerich,P.A.,  
 van der Groen,G., McCutchan,F.F.E., Eddy,G. and Burke,D.  
 AUTHORS Unpublished (1993)  
 TITLE Genetic diversity of the HIV-1 envelope glycoprotein  
 JOURNAL  
 STANDARD full automatic

#### COMMENT

These sequences were kindly provided prior to publication by Dr.  
 Joost Louwagie of the Henry M. Jackson Foundation Research Laboratory  
 Rockville, Maryland. Twenty-one full length gp160 coding sequences  
 from eight African countries (Djibouti, Gabon, Kenya, Senegal,  
 Somalia, Uganda, Zaire, and Zambia) were sequenced and analysed with  
 thirty-two previously published full-length gp160 env sequences.  
 Sequences DJ258, DJ259, K124, SE365, SM145, UC266, UC268, UC274,  
 V191, and V1925 have corresponding gag sequences published by  
 Louwagie et. al. in AIDS 7, 769-780 (1993) and in the 1993 compen-  
 dium. This sequence clusters with C subtype env sequences. See also  
 accession numbers L22939-L22957 and L23064.  
 NCBI gi: 437346

#### FEATURES

source 1..3340 Location/Qualifiers  
 /organism="Human immunodeficiency virus type 1"  
 /isolate="Djibouti"  
 /cell\_type="Lymphocyte"  
 /proviral  
 /sequenced\_mol="DNA"  
 join(87..>302,2619..>2707)  
 /note="tat protein, exon 2 (first expressed exon)"  
 /translation="MEPDPNLEPNWNGHSGSPKACTCYCKKSHCVLCFOTKGL  
 GISYKRRKRRGSRAPSSSEDHONLSKQ"  
 /note="tat protein, exon 3 (AA at 2620)"  
 /translation="PLSRQGDPTPEESKRVESKTRADPLD"

#### CDS



**CDS**

join(226..>302,2619...>2838)  
/note="rev protein, exon 2 (first expressed exon)"  
/translation="MAGRSGSDENLDAVRIKILYDS"  
/note="rev protein, exon 3 (AA at 2621)"  
/translation="SPYEPEKGTROAQNRNRRMRRAROIHSISERILSTCLGRPA  
EPVPOLPIERIENLDCSESGTSCTEGVN"  
325..>582

**CDS**

/note="vpu protein"  
/translation="MIDLAVVDRLVAALFIAFIATIIAYVTIAYIEIRKLROK  
IDRLIERIRERAEDSGNSDDTDDELSTMVDRGNLRLLDADV  
503...>3031  
/note="env polyprotein"  
/translation="MRVMGIOMNCQMMIGGILGFPMWLMCGNMGLMTWTYYGPV  
MOEAPRTFCASDAKAYETVEYNHNTACVPTDPSDELYEMNTENPNNKKMY  
DQMHDITSLMDGLPCVKLTPLCTYLNGSKRYATONGSYTISTEKDMNSFN  
ITELRDKRQRKRALFYKLDIVLPANGNEERLINCNSTIQAOPKVSFEPIAHY  
CAPGAYALKCRORVFENGTCPCKNVSTOCHTGIGLVYSTOLLNGSAEEDIIRS  
ENLNMAKIIIVOLNOSVEINCRPNNTROSIRIGPOTEFATDIIGDIRAOACH  
ISRCKMKETILOOVKGKLEHEFNKTIKFPASGGDIETIHSGNDGEFFCYNTSALF  
NDSTNPAINNSSDTIDATLCRKIOKLIIMNOEDREARAKYQCIITACKSNIGLL  
LTRRGFTSTDETFRPGGDMDNRKSLLIKYKEVAPLGIALPKARRVERKER  
AALAVALGFLGAAGSTWGANSITLTVAROLDSTIVQOSSLALKAIEHVHMQLT  
VMGKLOLQTRVALERYLKDOOLGWICMSCKLICTTPVPPNNSMSXSOEIMDNM  
TMWMODREINNTERIYRILEVSOEQONEDLADLKMOMLMPDITKMLAYIK  
IFINVGSLGILRIEFAIVSRVROGSYSFPTLLRNPRGPRPGIEEGEO  
DKGSTVRSGEFLAMDLRSLPEFSHRDLIYARVELLGGCWGETIKG  
SLVOYMGLELTKSAVSLDITTAIVAABSTDRIETITQIKMFACIPRRIRGFEAA  
LQ"  
3036..>3340  
/note="nef protein"  
/translation="MGNKWSGMPPAREIRIKTRPAAREVAAAAGVAAASODLKY  
GALTSTNPANNADCAMLPLEDEEEVEVFPRPVPLRPMTYKGAFLDFLKEG  
L"

**BASE COUNT** 1191 a 568 c 791 g 790 t

**ORIGIN** 226 bp upstream from the beginning of rev cds start

**Query Match** 3.0% Score 20; DB 3; Length 3340;  
**Best Local Similarity** 75.0%; Pred. No. 2; 6.9e-01;

**Matches** 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 570 tgcgttgatgctgaatggatggagaacttggcgac 609  
||||| ||||| ||||| || |||||  
58 TGCCTTGATGCTGATGCGATGGAGCTTTATATATC 97

**RESULT** 2 HIVDJ373A 3340 bp ds-DNA VRL 31-DEC-1993

**LOCUS** Human immunodeficiency virus type 1, DJ373 from Djibouti,  
**DEFINITION** proviral DNA encoding env, tat, vpu, rev, and nef genes.  
**ACCESSION** L22065  
**KEYWORDS** env gene; envelope glycoprotein; nef gene; rev gene; tat gene;  
vpu gene.

**SOURCE** Human immunodeficiency virus type 1 (Hiv-1), DJ273 from Djibouti,  
PCR amplification of proviral DNA from cocultivated PBMC's  
**ORGANISM** Human immunodeficiency virus type 1  
Viridae; SS-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
1 (bases 1 to 3340)

**REFERENCE** Louwagie,J.J., Janssens,M., Mascola,J.J., Fischer,C.L., Hegerich,P.A.,  
**AUTHORS** van der Groen,G., McCutchan,F.F.E., Eddy,G. and Burke,D.  
Genetic diversity of the HIV-1 envelope glycoprotein  
unpublished (1993)

**TITLE** full automatic  
**JOURNAL** These sequences were kindly provided prior to publication by Dr.  
**STANDARD** Joost Louwagie of the Henry M. Jackson Foundation Research Laboratory  
Rockville, Maryland. Twenty-one full length gp160 coding sequences  
from eight African countries (Djibouti, Gabon, Kenya, Senegal,  
Somalia, Uganda, Zaïre, and Zambia) were sequenced and analyzed with  
thirty-two previously full-length gp160 env sequences.  
Sequences D358, D3259, K124, SE365, SM45, UG366, UG368, UG374,  
VI191, and VI525 have corresponding gag sequences published by  
Louwagie et. al. in AIDS 7, 769-780 (1993) and in the 1993 compen-  
dium. This sequence clusters with C subtype env sequences. See also

Accession numbers U09876-U09905 and U09906.

NCBI gi: 437346

Location/Qualifiers

1..3340

/organism="Human immunodeficiency virus type 1"

/isolate="Djibouti"

/cell\_type="lymphocyte"

/proviral

/sequenced\_mol="DNA"

/tissue\_type="blood"

join(87..>302,2619..>2707)

/note="tat protein, exon 2 (first expressed exon)"

/translation="MEPVNDLPENMNSGQPTACRCKCCKKCSYCLVCFQTGLGISGRKKRRORRASPSSESDHONLSIKO"

/note="tat protein, exon 3 (aa at 2620)"

/translation="PLSRTOGDPPGPGESESKVESRKAPLDP"

join(226..>302,2619..>2838)

/note="rev protein, exon 2 (first expressed exon)"

/translation="MAGRSGDSDEALLGNRIIRIKLYQS"

/note="rev protein, exon 3 (aa at 2621)"

/translation="SPYPERGTRORAGNRRRARRARORIHISERILSTCLGRPAPVPEFPLPIERIENLDCSESGGTSGTEGVGN"

325..>582

/note="vpu protein"

/translation="MDLLAKVDYRLAVAFIAFIATIVMTIATIEYKKLRÖKIDRIERIEREAREDSNGESDGDDIELSTMVDRGNRLLDADV503..>9031

/note="env polyprotein"

/translation="MRVAGIONNCOQMIGILGFWMLLMGCMGNLMTVVYGVPMOEANPLPCASDAKAAYETEVHNYWATHACVPDPSPDELVMENVENEMKKNQYDOMMODIIISLMDEIGKCYKCTPLPCTVLINCNNVTARGNSSYTNSENREKRNCSTNITTELBRKÖKKRYALFKLDIVPLGENSEERYELINCTSSITÖACPRVSFEPIITHCAPGYAILKCRDTKEFGTGCPCKNVSTVOCTHIKPVYSTÖLLNGSTABEDIIRSCEANNKIIIVOLNOSVEINCTRPNNNTSIRIGGQTFYATGDIIGDIRÖHCNISROKMETLOOVYGKLKEHFHNKTIKFAPSGGIDETTHSFNGCGEFNCNSALFNDSSTNPINNNSNDANTITOCRIKOIIMNOVGAMVAMPYIOGIACKSNITGLLTRBGCNSTDETFRRPGGMRDNMWSBELKYVYEKPLGFLPAFRKRVVEREKAAALAVLFGLGAAGSTMGASTLTLYVAÖQLSLGIVÖÖSNLKAI EVHQHLÖLT VNGKÖLÖTVNLIAERYLKDÖQLSIGWGCKDLCTTAVVSNSSWSNKSOEEIWDN TMQMOEDREINNNYETIYRLLEVSÖTOEOENKELLDADKWNLSMSFDIÖKMYLIK IIMVÖGELIGLRIIFAVLIYNVRVGSYDFEÖTIPINPRPDPRRGREEGGÖDKGSTPVLGSGFLAAMDRLSLFEPHYHLRDLIIIVAVVSVILLÖRGHEFTLYIG SLVYOYGELEKKSANVSLDLTTATAIVAEGBTRIIEITIORIWRACFINPRIÖGFEAA LÖ"

CDS

-3036..>3340

/note="nef protein"

/translation="MGNKMSGMPAVFERIRTKRPAERYAAAEGVGAASÖLDXY GALTTSNTPANNADCANLEÖEEEBEGEVFRVQRVPLRPNTTYGARDLSFLKEKG L"

BASE COUNT

1191 a 568 c 791 g 790 t

ORIGIN

226 bp upstream from the beginning of rev cds start

Query Match

3.0% Score 20; DB 3; Length 3340;

Best Local Similarity 70.8%; Pred. No. 2.69e-01; Indels 0; Gaps 0;

Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Dc 1642 tacatcagcccttgittaatgattcaataccaacgaccaccaataataa 1689  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Cc 619 TAAATATGCACTTTTGTAATTGTAATTCACCICGACCCATAATATA 572

RESULT

LOCUS HIV251 303 bp ss-RNA VRL 09-SEP-1993  
DEFINITION Human immunodeficiency virus type 1, clone 2, ADC sample M, partial  
accession L21492  
SEGMENT 1 of 4  
ORGANISM Human immunodeficiency virus type 1 (HIV-1), clone 2, ADC sample M.  
SOURCE Human immunodeficiency virus type 1  
Virus: ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
REFERENCE  
1 (bases 1 to 303)  
Zhu,T., Mo,H., Wang,N., Nam,D.S., Cao,Y., Koup,R.A. and Ho,D.D.



TITLE Genotypic and phenotypic characterization of HIV-1 in patients with primary infection

JOURNAL Science 261, 1179-1181 (1993)

STANDARD full staff review

COMMENT This sample has been taken from patient M, one of the seven patients (A,M,L,F,C,V,R), in the study(1). All patients had acute, self-limited symptomatic illness with measurable viremia followed by seroconversion. Patient M was a chronically infected male who subsequently transmitted the virus to patient F. The authors report that the sequences in this study are found to be uniformly macrophage-tropic and non-synctium-inducing. The sequence for sample M clone 2, along with other patient M clones 1, 3-9, 11, and 12, is shown in alignment following the printed text entry. These gag sequences cluster with HIV-1 B subtype sequences. See also L21224-L21591 and L24161, L24162.

FEATURES

CDS

location/Qualifiers

<1..>303

/product="myristylated gag protein p17"

/gene="gag"

/codon\_start=1

/translation="LPSLQGTSEELSLNTVAVLCVHOKIDVKTKEALKEIEE QNKKKKAQAQATAADTRANSQVSNVPIVGNIGQWHQISPTILNWKVVEKA F"

1 (bases 1 to 309)

/organism="Human immunodeficiency virus type 1"

/sequenced\_mol="RNA"

128 a 64 c 62 g 55 t

BASE COUNT

128 a 64 c 62 g 55 t

source

128 a 64 c 62 g 55 t

BASE COUNT

128 a 64 c 62 g 55 t

Query Match

Best Local Similarity 77.1%; Pred. No. 1.15e+00;

Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 82 aagatagatgtaaaagacacacaaagacttaga 116

111 ||||| 111 ||||| 1111

170 AAGCTAGATGTTACAAATATTAAGAAGCATTTAGA 204

RESULT 4

LOCUS HIV10S1 309 bp ss-RNA VRL 09-SEP-1993

DEFINITION Human immunodeficiency virus type 1, clone 10, ADC sample F, partial gag cds, p17 region.

ACCESSION L21375

SEGMENT 1 of 4

SOURCE Human immunodeficiency virus type 1

ORGANISM Human immunodeficiency virus type 1 (HIV-1), clone 10, ADC sample F. Viridae; ss-RNA enveloped viruses; Positive strand RNA virus; Retroviridae; Lentivirinae.

REFERENCE 1 (bases 1 to 309)

AUTHORS Zhu,T., Mo,H., Wang,N., Nam,D.S., Cao,Y., Koup,R.A. and Ho,D.D.

TITLE Genotypic and phenotypic characterization of HIV-1 in patients with primary infection

STANDARD Science 261, 1179-1181 (1993)

COMMENT This sample has been taken from patient F, one of the seven patients (A,M,L,F,C,V,R), in the study(1). All patients had acute, self-limited symptomatic illness with measurable viremia followed by seroconversion. Patient F was infected by patient M via sexual transmission. The authors report that the sequences in this study are found to be uniformly macrophage-tropic and nonsynctium-inducing. The sequence for sample F clone 10, along with other patient F clones 1-9, 11, and 12, is shown in alignment following the printed text entry. These gag sequences cluster with HIV-1 B subtype sequences. See also L21224-L21591 and L24161, L24162.

FEATURES

CDS

location/Qualifiers

<1..>309

/product="myristylated gag protein p17"

/gene="gag"

/codon\_start=1

/translation="LPSLQGTSEELSLNTVAVLCVHOKIDVKTKEALKEIEE QNKKKKAQAQATAADTRANSQVSNVPIVGNIGQWHQISPTILNWKVVEKA F"

1 (bases 1 to 309)

/organism="Human immunodeficiency virus type 1"

/sequenced\_mol="RNA"

128 a 64 c 62 g 55 t

BASE COUNT

128 a 64 c 62 g 55 t

source

128 a 64 c 62 g 55 t

BASE COUNT

128 a 64 c 62 g 55 t

ORIGIN 190 bp downstream from beginning of gag cds

Query Match

Best Local Similarity 77.1%; Pred. No. 1.15e+00;

Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 82 aagatagatgtaaaagacacacaaagacttaga 116

111 ||||| 111 ||||| 1111

170 AAGCTAGATGTTACAAATATTAAGAAGCATTTAGA 204

RESULT 5

LOCUS HIV1415 1459 bp ss-RNA VRL 20-APR-1993

DEFINITION Human immunodeficiency virus type 1, isolate V1415 taken from a Rwandan national residing in Belgium, gag region.

ACCESSION L11791

SOURCE Human immunodeficiency virus type 1 (HIV-1), Rwandan isolate V1415.

ORGANISM Human immunodeficiency virus type 1 Viridae; ss-RNA enveloped viruses; Positive strand RNA virus; Retroviridae; Lentivirinae.

REFERENCE 1 (bases 1 to 1459)

AUTHORS Louwagie,J.J., McCutchan,F., Brennan,T., Peeters,M., Brennan,T., Sanders-Buell,E., Eddy,G., van der Groen,G., Fransen,K., Gershy-Damet,M., Deleys,R. and Burke,D.

TITLE Phylogenetic analysis of gag genes from seventy international HIV-1 isolates provides evidence for multiple genotypes

STANDARD AIDS 7, 769-780 (1993)

COMMENT Kindly provided prior to publication by Henry M. Jackson Foundation Research Laboratory, 1500 East Gude Drive, Rockville, MD 20850. The V1415 gag sequence clusters with HIV-1 A subtype sequences.

FEATURES

CDS

location/Qualifiers

1..1459

/product="gag protein"

/gene="gag"

/codon\_start=1

/translation="MGARASVLSGKLDAMEKIRLRPGRRKRYRMKHLVNASRELDRE ALNPGLETAEGCOOILEOLPALKTGEIKSLVNTVATLVCHORIDVCKREALD KIEIRKRNKQKTDQAAAGTGNSSNVSQNPYONAGOMTHQASIRPTLNARVAYE EKAPSEVPIPFESALSGATPQDINMLNLTGVGQAOAMKIDTINEAEWDLHFV HAGPIPGOMKREPGRSDIAGTSTQDQIMKGNPNPVGDIYKRIIIGLNKIVM VSPSIDVLEKGPKEPRDYVDFEFKILRAEOATOEYGMWTEILVONANPDCKTIL RALGTGATLEEMVACOGVGPGKRAVLAEASVOHTNIMMORNGFKOORIKCN CGREGHARNCRAPRRKCGKCKCKGCKGMDCTERONPLFKITSSSKGPGNPGSR PEPTAPPAEIVMGGEISPPKQREDAPELVLSKSX"

1..1459

/organism="Human immunodeficiency virus type 1"

/isolate="V1415"

/cell\_type="lymphocyte"

/proviral

/sequenced\_mol="DNA"

/tissue\_type="blood"

BASE COUNT 542 a 278 c 366 g 273 t

ORIGIN gag cds start

Query Match

Best Local Similarity 75.7%; Pred. No. 1.15e+00;

Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 269 agagatagatgtaaaagacacacaaagacttaga 305

111 ||||| 111 ||||| 1111

168 AGAAGCTAGATGTTACAAATATTAAGAAGCATTTAGA 204

RESULT 6

LOCUS SIYMDG31 9215 bp ss-RNA UNA 20-DEC-1989

DEFINITION Simian immunodeficiency virus from a mandrill, isolate GBL.

ACCESSION M27470.115781

REFERENCE 1 (bases 1 to 9215)

AUTHORS Tsujimoto,H., Hasegawa,A., Maki,N., Fukasawa,M., Miura,T., Speidel,S., Cooper,R.W., Moriyama,E.N., Gojobori,T. and Hayami,M.

TITLE Sequence of a novel simian immunodeficiency virus from a wild-caught African mandrill



JOURNAL  
STANDARD full staff review  
COMMENT The mandrill virus is distinct from all other primate immuno-deficiency viruses, thus it can be regarded as a type 4 virus. There is neither a vpx nor a vpu coding sequence. The splice sites and coding regions for tat and rev are tentative.  
location/Qualifiers  
CDS <1745..4774  
/note="pol (NH2-terminus uncertain; AA at 1745)"  
/codon\_start=1745  
join(5471..5730,7950..8037)  
/note="tat protein"  
/codon\_start=5471  
join(5590..5677,7950..8167)  
/note="rev protein"  
/codon\_start=5590  
275..291  
misc\_binding  
/note="primer (lys-trna) binding site"  
450..1958  
/note="gag polyprotein"  
/codon\_start=450  
4728..5246  
/note="vif"  
/codon\_start=4728  
5227..5541  
/note="vpr protein"  
/codon\_start=5227  
5471..5730  
exon  
/note="tat protein, exon 2 (first expressed exon)"  
5390..5677  
exon  
/note="rev protein, exon 2 (first expressed exon)"  
5661..8126  
CDS  
/note="env polyprotein"  
/codon\_start=5661  
7950..8037  
exon  
/note="tat protein, exon 3 (AA at 7951)"  
7950..8167  
exon  
/note="rev protein, exon 3 (AA at 7952)"  
8170..8814  
CDS  
/note="nef"  
/codon\_start=8170  
9191..9196  
misc\_signal  
/note="poly-A signal"  
9191..9196  
BASE COUNT 3323 a 1478 c 2196 g 2218 t  
ORIGIN

Query Match 2.7%; Score 18; DB 2; Length 9215;  
Best Local Similarity 90.9%; Pred. No. 4.46e+00;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 485 ttggataatttgagaata 506  
||||| ||||| ||||| |||||  
QY 446 TTGGATCAATTAGAGAAATA 467

RESULT 7  
LOCUS HIVLAI 9229 bp ss-RNA VRL 01-MAR-1989  
DEFINITION Human immunodeficiency virus type 1, isolate LAI, complete  
ACCESSION K02013 M64186-190 M64194-198 M64203-223  
KEYWORDS acquired immune deficiency syndrome; complete genome;  
env gene; gag gene; long terminal repeat; pol gene; polyprotein;  
provirus; reverse transcriptase; trans-activator;  
Human immunodeficiency virus type 1 (HIV-1), isolate LAI proviral  
DNA clone lambda-119 (Infectious clone available; formerly HIVBRU).  
ORGANISM Human immunodeficiency virus type 1  
Virus: ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.  
REFERENCE 1 (bases 1 to 9229)  
AUTHORS Main-Hobson, S., Sonigo, P., Danos, O., Cole, S. and Alizon, M.  
TITLE Nucleotide sequence of the AIDS virus, LAV  
JOURNAL Cell 40, 9-17 (1985)

STANDARD full staff review  
REFERENCE 2 (bases 1712 to 1749; correction of [1] in gag-pol cds)  
AUTHORS Alizon, M., Main-Hobson, S., Montagnier, L. and Sonigo, P.  
TITLE Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isolates from African patients  
JOURNAL Cell 46, 63-74 (1986)  
STANDARD full staff review  
AUTHORS  
REFERENCE 3 (gag, tat, rev, env V1 and env V3 regions of LAI samples)  
Main-Hobson, S., Vartanian, J.-P., Henry, M., Chenciner, N., Cheynier, R., Delassus, S., Pedrozza Martins, L., Sala, M., Nngeye, M.-T., Guetard, D., Barre-Sinoussi, F. and Montagnier, L.  
LAV revisited: origins of the early viral HIV-1 isolates from Institut Pasteur  
Science 252, 961-964 (1991)

JOURNAL  
STANDARD  
COMMENT The original LAV, sometimes called LAV-1 to distinguish it from HIV2(LAV-2) and formerly designated HIVBRU, is now designated HIVLAI to reflect the finding that it originated with the French AIDS patient LAI (and not the patient BRU) as explained in [3]. An infectious clone of this virus has been constructed by Keith Peden, NIH-NIAID, Bethesda, 301-496-6730. HIVLAI3 is also an infectious clone having for its 3' half a clone of the LAI isolate. The tat sequence for the L20 clone is available (see 1989: I-A-34 and following pages in this release drawing upon PCR-generated sequence sets [3]). All LAV, BRU and IIB viral samples are now believed to have been derived from this source, LAI [3]; viruses from the person BRU are now designated HIVJB.  
location/Qualifiers  
FEATURES  
LTR <1..180  
/note="5' LTR"  
<1..97  
/note="R repeat 5' copy"  
<1631..4678  
/note="pol polyprotein (NH2-terminus uncertain; AA at 1631)"  
/codon\_start=1631  
join(5412..5626,7972..8017)  
/note="tat protein"  
/codon\_start=5412  
join(5551..5626,7972..8246)  
/note="rev protein"  
/codon\_start=5551  
prim\_transcript 1..9229  
/note="genomic mRNA"  
prim\_transcript 1..9229  
/note="tat, rev, nef subgenomic mRNA"  
182..199  
/note="primer (lys-trna) binding site"  
290..5358  
/note="tat, rev, nef subgenomic mRNA Intron 1"  
336..1874  
/note="gag polyprotein"  
/codon\_start=336  
/location=Cell 40, 9-17 (1985) | 1712..1713  
4623..5201  
/note="vif protein"  
/codon\_start=4623  
5141..5431  
/note="vpr protein"  
/codon\_start=5141  
5412..5626  
/note="tat protein, exon 2 (first expressed exon)"  
5551..5626  
/note="rev protein, exon 2 (first expressed exon)"  
5627..7971  
/note="tat cds intron 2"  
5627..7971  
/note="rev cds intron 2"  
5627..7971  
intron  
/note="tat, rev, nef subgenomic mRNA Intron 2"  
5643..5888  
CDS  
/note="vpu protein"



```

CDS      /codon_start=5643
         /note="envelope polypeptide"
         /codon_start=5803
         5803..8388
         7972..8017
         /note="tat protein, exon 3 (AA at 7973)"
         7972..8246
         /note="rev protein, exon 3 (AA at 7974)"
         8390..9010
         /note="nef protein"
         /codon_start=8390
         8679..9229
         /note="3' LTR"
         9133..9229
         repeat_region
         /note="R repeat 3' copy"
         9205..9210
         /note="mRNA polyadenylation signal"
         misc_signal
         3289 a 1656 c 2232 g 2052 t
         ORIGIN
         Cap site of genomic RNA.

Query Match
Best Local Similarity 78.1%; Pred. No. 4.46e+00;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 3393 cttagtgaattatgtaccagtagagaaa 3424
   ||| ||||| ||| ||| ||||| |||
   433 CTTATGTGAATTTTGGATCAATTAGAGAAA 464

RESULT 8
LOCUS    HIVBRU      9229 bp ss-RNA      VRL      01-MAR-1989
DEFINITION Human immunodeficiency virus type 1, isolate BRU, complete
            genome (original LAV-1)
ACCESSION K02013
SOURCE    Human immunodeficiency virus type 1 (HIV-1), isolate BRU
            proviral DNA clone lambda-919 (infectious clone available).
REFERENCE 1 (bases 1 to 9229)
AUTHORS   Main-Hobson,S., Sonigo,P., Danos,O., Cole,S. and Alizon,M.
TITLE     Nucleotide sequence of the AIDS virus, LAV
JOURNAL   Cell 40, 9-17 (1985)
STANDARD
REFERENCE 2 (bases 1713 to 1748; revision of [1])
AUTHORS   Alizon,M., Main-Hobson,S., Montagnier,L. and Sonigo,P.
TITLE     Genetic variability of the AIDS virus: Nucleotide sequence analysis
            of two isolates from African patients
JOURNAL   Cell 46, 63-74 (1986)
STANDARD
COMMENT   The original LAV, sometimes called LAV-1 to distinguish it from
            HIV2 (LAV-2), is now referred to as HIV-1brU. An infectious
            clone of this virus has been constructed by Keith Peden, Molecular
            Biology and Genetics, Johns Hopkins University School of Medicine,
            Baltimore, MD 21205 (301) 955-3652. HIVNL43 is also an infectious
            clone having for its 3' half a clone of the BRU isolate. The tat
            sequence for the L20 clone is available (see 1989: I-A-54).
FEATURES
pept     < 1631 1874 4678
         /note="gag polypeptide"
         /note="pol polypeptide (NH2-terminus uncertain; AA at
         1631)"
         5201
         /note="vif protein"
         5141
         /note="vpr protein"
         5412
         /note="tat protein, exon 2 (first expressed exon)"
         7972
         /note="tat protein, exon 3 (AA at 7973)"
         5551
         /note="rev protein, exon 2 (first expressed exon)"
         7972
         /note="rev protein, exon 3 (AA at 7974)"
         5643
         /note="vpu protein"
         5888
         /note="env polypeptide"
         8388
         /note="nef protein"
         9010
         /note="tat, rev, nef subgenomic mRNA intron 1"
         5627
         /note="tat cds intron 2"
         7971
         /note="rev cds intron 2"
         180
         /note="3' LTR"
         8679
         /note="5' LTR"
         9229
         /note="R repeat 5' copy"
         rpt < 1 97

```

```

rpt      9133 9229 R repeat 3' copy
binding  182 199 primer (Lys-tRNA) binding site
reftnumb 1 1 numbered 1 in [1]
revision 1712 1749 attcttcagagcagacagagccacagcccccacag in [2];
          ag in [1]
signal   9205 9210 mRNA polyadenylation signal
BASE COUNT 3289 a 1656 c 2232 g 2052 t
ORIGIN    Cap site of genomic RNA.

Query Match
Best Local Similarity 78.1%; Pred. No. 4.46e+00;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 3393 cttagtgaattatgtaccagtagagaaa 3424
   ||| ||||| ||| ||| ||||| |||
   433 CTTATGTGAATTTTGGATCAATTAGAGAAA 464

RESULT 9
LOCUS    SIVAGM155  9794 bp ds-DNA      VRL      14-NOV-1989
DEFINITION Simian immunodeficiency virus from African Green Monkey,
            isolate 155, clone 4; complete genome.
ACCESSION M29975
SOURCE    Simian immunodeficiency virus, isolate 155, clone 4; proviral DNA.
            Biologically active clone.
REFERENCE 1 (bases 1 to 9794)
AUTHORS   Johnson,P.R., Fomsgaard,A., Allan,J., Gravel,M., London,W.T.,
            Olmstead,R.A. and Hirsch,V.M.
TITLE     Simian immunodeficiency viruses from African green monkeys display
            unusual genetic diversity
JOURNAL   J. Virol. 64, 1086-92 (1990)
STANDARD
COMMENT   full staff_review

Kindly submitted prior to publication and in a computer readable
form by Phillip Johnson, Georgetown University, Rockville MD (301-
496-2976). The 155 isolate is from a monkey imported from Kenya.
FEATURES
CDS      join(6051..6268,8492..8633)
         /note="tat protein"
         /codon_start=6051
         join(6208..6268,8492..8700)
         /note="rev protein"
         /codon_start=6208
         1..726
         /note="5' LTR"
         1..625
         /note="R repeat 5' copy"
         727..744
         /note="primer (Lys-tRNA) binding site"
         931..2493
         /note="gag polypeptide"
         /codon_start=931
         2199..5342
         /note="pol polypeptide"
         /codon_start=2199
         5260..5958
         /note="vif protein"
         /codon_start=5260
         5741..6100
         /note="vpx protein"
         /codon_start=5741
         6051..6268
         /note="tat protein, exon 2 (first expressed exon)"
         6208..6268
         /note="rev protein, exon 2 (first expressed exon)"
         6275..8581
         /note="env polypeptide"
         /codon_start=6275
         8492..8633
         /note="tat protein, exon 3 (AA at 8493)"
         8492..8700
         /note="rev protein, exon 3 (AA at 8494)"
         8724..9416
         CDS

```



```
/note="nef protein"
/codon_start=8724
LTR
9070..9794
/note="3' LTR"
repeat_region
9578..9794
/note="R repeat 3' copy"
misc_signal
9675..9680
/note="poly-A signal"
BASE COUNT
3321 a 1905 c 2450 g 2118 t
ORIGIN
Query Match
2.7%; Score 18; DB 3; Length 9794;
Best Local Similarity 78.1%; Pred. No. 4,46e+00;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 5377 tacaagaatcattcaatcaacttgagcattg 5408
OY 603 TACAAATGCATATTAGATAATTAGGCATG 634

RESULT 10
LOCUS HYTH475A 9795 bp ss-RNA VRL 22-Apr-1994
DEFINITION Human immunodeficiency virus type 1, isolate TH475A, complete
genome.
ACCESSION L31963
KEYWORDS
SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate TH475A.
ORGANISM Human immunodeficiency virus type 1
Virus: ss-RNA enveloped viruses; Positive strand RNA virus;
Retroviridae; Lentivirinae.
1 (bases 1 to 9795)
REFERENCE
1 Neumann,M.W.F., Kilschmidt,A., Felber,B.K., Froese,B., Ertel,Y.,
Paylakis,G.N. and Brack-Werner,R.
Association of HIV-1 production in human astrocytoma cells is
associated with a cellular block in Rev function
Unpublished (1994)
TITLE
HIVTH475A> clusters with subtype B viruses.
NCBI gi: 474287
FEATURES
location/Qualifiers
1..9795
/organism="Human immunodeficiency virus type 1"
/isolate="TH4-7-5"
/strain="111b"
/proviral
/sequenced_mol="RNA"
828..>2327
/note="gag polyprotein"
/translation="MGARASYLSGGELDMREKIRLRPGKKKKYKLIHIVASRELER
FAVNPGLATSKGROILGOLPSLQSGSEERSLYNTVATLYCVHGRIRIKDTKEA
LDKIEEONKSKKAAQAAADTGHSSOSONYPIVONIGOMYQAOAISPTLNAVY
VVEEKASPEVIMWESALSAGATPDOLNTMLNTYGGHQAOMKKEINDEAAEMDR
VHVPHAGPIAPGOMREPRGSDIAGTSTLDBQIGMNNPPIPEGLIKKWTLLGLN
KITYRMSPTSLIDLRGPKPEPRDYVIRFKTLIAEQASQEVKNMWTETLLVONAP
DCETIKALGPATLEEMMTACQGVGPKARVLAAMSQVNSATIMQORERN
QRKIVCFNCGEGHTARNCRAPRKCKMCKGEGHQMCKCTEROANFLFKIPDSYK
GRPNFLQSRPEPTAPPEESFRSGVETTPPOKEPIDKELYPLTSLRSFLGNDPS
0"
CDS
<2123..>5128
/note="pol polyprotein (AA at 1)"
/translation="FFREDLAFLOGKARESSSDOTRANSPTRRELOWGHDNNSPSE
AGADROGVSNFPOITLMORPLVTIKIGOLKEALDITGADTVLEMSLPRMPP
KMIGIGCFIKVRODOIIEICGHKAIQVLPVPPVNTIGRNLLOIGCTLNPI
SPTEIVPVKLPKPGMDGPKVOMPLTEEKIATVIEICMEMEKEGKISIGENYNP
VFAIKKDSKMKKLVDFRELNRKTDQDFWEQIGIPRALKKKSTTVLDVDATE
SVLDEDFRKYTAFTPSINNEPGRITQYNVLPQKKGSPALFQSSMTILDEPRK
QNDPIVYQYMDLYVGSDEIGQHRKIELEKQHLRMGFTLPDKKHQEPFLM
GVELHPDKMTVOPVLPLEKDSVTYNDIOKLVGLNMAQSPICIKVLOCLKLGRK
ALTEVLPTEAELELAENREILKEPVHGVLYNDPSKLIIEIOKGGOWTLYIOYE
PFNKLTGKAYARMGANTNDVKOLEAVOKITTESIYIMCKEKKFKIPLOKEWEM
WTEYKQATPIPEPEFVNTPLPVKIMYOLEKEPVLGATTFYVDGANREIKLGRAGY
TNGRQKVVNTIDTNTOKTELQALYLVLOSSEVNTVDSOALGIIQAPKSES
ELVNOIIEOLIKREKITYLAWPAHKIGGNEOVKILVSAGIRVPLFDIGDKADEH
EKHSNMRAMASDFNLPPVAKETIVASCRCOLKGEAMGQVDCSPGIWOLDCTHE
GKYLIVAVHVASGYIEAVYPAETGOETAEFLIKLGRMPVKTIDTNSNFTSATV
KACMWMAGIKOEKGIPIYNPOSOVVSAMKELKIIIGOVRODAEHKTAQVAVFIH
NKRKGIGIGISAGERIVDIATIDTQIKELQKQITKIQNRVYIRDSRDLMPGPAK
LWKGGGAVVYIQNSDIKVPARRKARIIRYQKQMGADDCVARSODE"
5079..>5655
/note="vif protein"
/translation="MENRQVMYIMQVDMRIRITKMSLYKHHMYSGKARGMFRRH
YESPRPISSEVHPIGDARIVTITWGTIRGDRPDHLOQGVSEIRKRRYSQVDP
ELADQILHLIYFPCFSDSARIRKALLGHYSPREYQAGHKVGSLOYLDLALITPK
KIRPLPSVTKTLEDKRNKPKQTKGHRGSHTNH"
5597..>5708
/note="vpr protein (premature termination at 5708)"
/translation="MEQAPDQGPQREPHNMTLELLELKNVAVHPPRI"
5597..>5885
/pseudo
/note="vpr protein"
join(5868..>6082,8416..>8460)
/note="tat protein, exon 2 (first expressed exon)"
/translation="MEVPDRLPEPKMKHPSGSPKTAICTCYCKKCFHCQCFMTKAL
GISYGRKKRORRRARHNSOTHOASLSKQ"
/note="tat protein, exon 3 (AA at 8454)"
/translation="PTSQSGDPTGKE"
join(6007..>6082,8416..>8689)
/note="rev protein, exon 2 (first expressed exon)"
/translation="MAGRSGSDDELIRVRLIKLYQS"
/note="rev protein, exon 3 (AA at 8455)"
/translation="NPPNPEGRORARRRRRRRRRQOIHISIRILSTYLGRA
EVPDLPLERLTLDNCEDCGISGIVGSPQIVLESPTILSGAKE"
6099..>6345
/note="vpu protein"
/translation="MOPPIVIVALVVAIIIAIVVSIIVIEYRKILRQKIDRLI
DLIERAEDSGNSESEISALVEMVGMHNAFWDIDL"
6262..>8829
/note="envelope polyprotein"
/translation="MRVKEKYOLHMRMGKMGMTLGLIMICSATEKLVWYIYGV
VWKEATITLFCASDAKAYDTEVHNWATACVTDNDPEVYLVNTEFNMMKNM
VEQMHDIITLSCADSLKPCVKLPLCVSLKCTDKLDKNTNNSGSMIEKEIKNC
SENISIKRGKQKEAYAFYKLDIPIINDTTSYPTSCNTSVITQACRVSPEPI
IHYCAGAPAILKCNKNTFNGPCPNVSTVOCTHRIPIVYSQILNLSLAEVY
IKSVNTDIAKTIIVOLNTSEINCRIPNNKTRKRIORGPERAVNTIGKIGNMRO
AHGNTISRAKMMNATLQIASKLREQFNNTIIFKQSGSDPEIYVHSFGGEFEYC
NSYOLNENSTWNSNTLSTEGSNNTEGSDITLPCRIOIIMMQEVGKAYAPISGO
IRCSSNITGLLDRDGSNNSESEIPRPGGMDRMRSLEYKVKYKTEPIGVAPT
KAKRRVOREKRAVYGALFIFGLFISAGSTMAASMTLVQAROLLSGIVQOONNL
RATEAOHLIQLTWGIRKODARILAVEYKLDQDLIGTSGSKICTTAPYRNAS
WSMKSLEQIMNHTMMDREINNTYSLHSLDESOQOENNEDELDELFWASLW
MPEDISSWLWYIKIFIMIVGGLVGLRIVAVSVVNRQGVSPISFQTHLPIDP
DRREGIEEGGERDSDRSIRLVNGLSLALIMDLRSICLFSYRRLDLILIVRIEY
LGRGWEALAKYMMNLQOYVSOELKNSAVSLNATAIYAVGTDVIEVQAVRAIR
LHRRIRKQELERTLL"
8834..>9452
/note="nef protein"
/translation="MSGKMSKSSVIGMPAVRERMRARRAEPADGVAGVSHLEKGA
TSSNTANNAACAMLEQDEEVEGFPVTOVLRPMYTAADLSHLEKKGGLGL
HSQRDLIDLMYITQCYFPDMQVYTGPGIRAPLFGMCKYKLVAVPEPLEBAN
KGNSTSLHPVSLHGMDDPEREYLERPDRSLAFHNVAVNELRPEYKNC"
<1..>782
/note="5' LTR (AA at 2)"
9123..>9617
/note="3' LTR"
repeat_region
463..>560
/note="R repeat 5' copy"
BASE COUNT
3436 a 1782 c 2389 g 2188 t
ORIGIN
Query Match
2.7%; Score 18; DB 3; Length 9795;
Best Local Similarity 78.1%; Pred. No. 4,46e+00;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 3849 ctttagtgaattatggtacagtagagaaa 3880
OY 433 CTTATGGAATTTTGGATCAATTAGAGAAA 464
```



**RESULT 11**  
**LOCUS** 1882 bp ss-RNA  
**DEFINITION** Simian (stump-tailed macaque) immunodeficiency virus, isolate STM,  
 complete genome, proviral DNA.  
**ACCESSION** PENDING  
**SOURCE** Simian (stump-tailed macaque) immunodeficiency virus, isolate STM,  
 bioactive, PCR-derived clone 37.16, proviral DNA.  
**REFERENCE** 1 (bases 1 to 9892)  
**AUTHORS** Johnson, P. and et.al.  
**TITLE** Unpublished  
**JOURNAL** Personal communication (1991)  
**STANDARD**  
**REFERENCE** 2 (sites)  
**AUTHORS** Novembre, F.J., Hirsch, V.M., McClure, H.M., Fultz, P.N. and Johnson, P.  
**TITLE** SIV from stump-tailed macaques: molecular characterization of a  
 highly transmissible primate lentivirus  
**JOURNAL** Virology (1991) in press  
**STANDARD**  
**COMMENT** The 5' LTR contains a deletion in U3 (possible PCR artifact) that  
 does not affect biologic activity. The 3' LTR is not full-length  
 due to placement of the PCR primers.  
**FEATURES**  
 from to/span description  
 pept 709 2232 gag polyprotein (AA at 1932; NH-2 terminus  
 pept < 1932 5057 pol polyprotein  
 pept 4987 5631 vif protein  
 pept 5459 5787 vpx protein  
 pept 5799 6104 vpr protein  
 pept 5950 6239 tat protein, exon 2 (first expressed exon)  
 pept 8454 8553 tat protein, exon 3 (AA at 8454)  
 pept 6176 6239 rev protein, exon 3 (first expressed exon)  
 pept 8454 8692 rev protein, exon 3 (AA at 8455)  
 pept 6246 8894 envelope polyprotein  
 pept 8728 9522 nef protein  
 retnumbr 1 1 numbered 1 in [1]  
 LTR 1 482 5' LTR  
 LTR 9113 > 9892 3' LTR  
 rpt 191 367 R repeat 5' copy  
 rpt 9633 9809 R repeat 3' copy  
 binding 486 504 primer (Uys-tRNA) binding site  
 signal 9786 9791 mRNA polyadenylation signal  
**BASE COUNT** 3315 a 1880 c 2529 g 2168 t  
**ORIGIN** 5' LTR start  
 Query Match 2.7%; Score 18; DB 3; Length 9892;  
 Best Local Similarity 68.8%; Pred. No. 4.46e+00;  
 Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

**FEATURES**  
**CDS** Location/Qualifiers  
 <1..>276  
**BASE COUNT** 120 a 43 c 47 g 66 t  
**ORIGIN** 811 bp downstream from the beginning of env cds  
 Query Match 2.6%; Score 17; DB 3; Length 276;  
 Best Local Similarity 84.0%; Pred. No. 1.55e+01;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

**Db 44** taatgtacactgtgaagcgt 68  
 || ||| ||||| ||||| |||||  
**QY 543** TATTAGAACATTGATTAATGCGTGT 567

**RESULT 13**  
**LOCUS** HIV008636 276 bp ss-RNA  
**DEFINITION** Human immunodeficiency virus type 1, sample 016 clone D from  
 Rwanda, C2V3 of env cds.  
**ACCESSION** U08636  
**SOURCE** Human immunodeficiency virus type 1 (HIV-1), sample 016 clone D  
 from Rwanda; derived from a primary isolate.  
**REFERENCE** 1 (bases 1 to 276)  
**AUTHORS** WHO Global Programme on AIDS.  
**COMMENT** This sample is part of a set of sequences generated through the WHO  
 Global Programme on AIDS. The virus was derived from an  
 asymptomatic individual, from Rwanda, whose route of infection is  
 thought to be due to heterosexual contact. The blood sample was  
 taken in 1992. This env sequence clusters with HIV-1 A subtype  
 sequences. The full name of this sequence is  
 HIV192RW016WHO.01.D1GCR; it was presented in alignments in an  
 abbreviated form in the April 94 Human Retroviruses and AIDS  
 compendium update as A2RW016W.01.D1GCR.  
**FEATURES**  
**CDS** Location/Qualifiers  
 <1..>276  
**BASE COUNT** 119 a 45 c 48 g 64 t  
**ORIGIN** 811 bp downstream from the beginning of env cds  
 Query Match 2.6%; Score 17; DB 3; Length 276;  
 Best Local Similarity 84.0%; Pred. No. 1.55e+01;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

**Db 44** taatgtacactgtgaagcgt 68  
 || ||| ||||| ||||| |||||  
**QY 543** TATTAGAACATTGATTAATGCGTGT 567

**RESULT 14**  
**LOCUS** HIV004917 314 bp ds-DNA  
**DEFINITION** Human immunodeficiency virus type 1, isolate 712 from  
 Baltimore, U.S.; partial env cds, C2V3 region.  
**ACCESSION** U04917  
**SOURCE** Human immunodeficiency virus type 1 (HIV-1), isolate 712,  
 Baltimore, U.S.; cultured on PBMCs.  
**REFERENCE** 1 (bases 1 to 314)  
**AUTHORS** NIAID/NIH DAIDS Variation Program.  
**COMMENT** This sample is part of a set of sequences generated through the  
 NIAID/NIH DAIDS HIV variation program. The virus was derived from  
 an asymptomatic individual, from Baltimore, U.S., whose route of  
 infection is thought to be due to heterosexual contact. The  
 blood sample was taken in 1992. This env sequence clusters with  
 HIV-1 B subtype sequences. The full name of this sequence is  
 HIV192US712DAI8A.01.D1LSDD; it was presented in alignments in an  
 abbreviated form in the April 94 Human Retro. AIDS compendium  
 update as B2US712D.01.D1GCR.  
**FEATURES**  
**CDS** Location/Qualifiers  
 <1..>314  
**BASE COUNT** 139 a 43 c 55 g 77 t  
**ORIGIN** 835 bp downstream from beginning of env cds  
 Query Match 2.6%; Score 17; DB 3; Length 314;



Best Local Similarity 79.3%; Pred. No. 1.55e+01;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 209 ctataaattaaagaacatttgagaat 237  
||||| ||| ||||| ||||| |||||  
QY 534 CTATTAATGATTAAGAACATTTGATTAAT 562

## RESULT 15

LOCUS HIVM12199 334 bp ds-DNA VRL 03-JUN-1993  
DEFINITION Human immunodeficiency virus type 1, sample M12199 from Malawi,  
partial env cds, V3 region.

## ACCESSION

L15721

## SOURCE

Human immunodeficiency virus type 1 (HIV-1), Malawi sample 12199.

## ORGANISM

Human immunodeficiency virus type 1  
Virus: ss-RNA enveloped viruses; Positive strand RNA virus;  
Retroviridae; Lentivirinae.

## REFERENCE

1 (bases 1 to 334)  
Orloff,G.M., Kalish,M.L., Chiphangwi,J., Potts,K.E., Ou,C.,  
Schochetman,G., Dallabetta,G., Saah,A.I. and Mottl,P.G.

## TITLE

V3 loops of HIV-1 specimens from pregnant women in Malawi uniformly  
lack a potential N-linked glycosylation site

## JOURNAL

AIDS Res. Hum. Retroviruses 9, 705-706, 1993

## STANDARD

full staff review

## COMMENT

Partial HIV-1 envelope sequences obtained from PCR amplified  
PBMC-cell DNA of infected pregnant women enrolled in a study  
of risk factors for HIV-1 in Blantyre, Malawi. An alignment of  
the sequences in this set (accession numbers L15721-L15735),  
follows the printed text entry. All of the sequences in this  
study belong to the env subtype C.

## FEATURES

Location/Qualifiers

## CDS

&lt;1..&gt;334

## source

/standard\_name="gp120"  
/product="envelope glycoprotein"  
/gene="env"  
/note="putative"

## V\_region

1..334

## source

/standard\_name="V3-loop"  
/gene="env"  
/organism="Human immunodeficiency virus type 1"  
/isolate="Malawi patient 12199"  
/proviral  
/sequenced\_mol="DNA"

## BASE COUNT

146 a 53 c 59 g 76 t

## ORIGIN

826 bp downstream from the beginning of env cds

## Query Match

2.6%; Score 17; DB 3; Length 334;  
Best Local Similarity 79.3%; Pred. No. 1.55e+01;

## Matches

23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

## Db

233 atttctaatatacaagaatttga 261

## Cp

554 ATTGTCTAATACATTAATGAATCTCA 526







## Search completed:

Sat Nov 27 12:36:00 1999

## Job time :

9 secs.









Db 687 attaaagattcggatgacattatgctcaaaaatactcaattagcgataaagtctgc 746  
Oy 1 atttaaacattcggatgacattatgctcaaaattttcaatttcggataaagctgc 60  
Db 747 ttctgatgtgatgtgatgtgagtttatgatcactctctctcaacaatagtga 806

RESULT 4  
ID V53479 standard; DNA: 619 BP.  
AC V53479;  
DE 30-OCT-1998 (first entry)  
DT DNA encoding a Staphylococcus aureus protein of unknown function.  
KW Staphylococcus aureus protein; immune response induction; eye infection;  
KW antibody production; T-cell immune response; gastrointestinal infection;  
KW respiratory infection; inhibitor; bacterial infection; cardiac infection;  
KW central nervous system; kidney infection; urinary tract infection;  
KW antimicrobial compound identification; broad spectrum antibiotic;  
therapy; ss.  
OS Staphylococcus aureus.  
PN EP-841394-A2.  
PD 13-MAY-1998.  
PF 24-SEP-1997; 307485.  
PR 24-SEP-1996; US-027032.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PI Black MT, Burnham MKR, Hodgson JE, Knowles DJC,  
PI Lometto KM, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,  
PI Ward JM;  
DR WPI, 98-252940/23.  
DR P-PSDB: W77686.  
PT New nucleic acid sequences from Staphylococcus aureus WCHU29 -  
PT useful in vaccines and for treatment of bacterial infections of e.g.  
PT respiratory tract and central nervous system  
PS Claim 1; Page 146; 390pp; English.  
CC This sequence encodes a Staphylococcus aureus protein of unknown  
CC function, and represents a DNA sequence of the invention.  
CC The DNA sequences were isolated from Staphylococcus aureus WCHU29  
CC (NCIMB 40771). Host cells containing the DNA sequences are used to  
CC produce polypeptides or fragments. The proteins are used in the treatment  
CC of disease, for inducing an immune response by administering them, to  
CC produce antibody and/or T-cell immune response. Antagonists of the



Query Match	88.98;	Score 587;	DB 48;	Length 619;
Best Local Similarity	98.88;	Pred. No. 0.00e+00;		
Matches 599;	Conservative 0;	Mismatches 6;	Indels 1;	Gaps 1

DR	WP1: 99-045171/04.
PT	New isolated <i>Enterococcus faecalis</i> polynucleotides and polypeptides
PT	- used to develop products for the detection of <i>Enterococcus</i> and for
PT	use in vaccines for prevention or attenuation of <i>Enterococcus</i>
PT	infection.
PS	Claim 1: Page 1983: 2084pp: English.
CC	A computer readable medium has been developed which has recorded on it
CC	982 nucleotide sequences isolated from the <i>Enterococcus faecalis</i> genome.
CC	X12938 to X131919 represent these nucleotide sequences which are primary
CC	nucleotide sequences, also known as contigs. The computer-based system
CC	can identify fragments of the <i>Enterococcus faecalis</i> genome with
CC	commercial importance. The products can be used to detect the presence
CC	of <i>Enterococcus faecalis</i> in samples. They can also be used for
CC	diagnosing Enterococcal infection in an animal and monitoring
CC	progression of disease, and for identifying agents which can be used to
CC	modulate the growth or pathogenicity of <i>Enterococcus faecalis</i> , or
CC	another related organism, in vivo or in vitro. In particular the
CC	polypeptides encoded by the <i>Enterococcus faecalis</i> nucleotide sequences
CC	can be used in vaccines to prevent or attenuate an Enterococcal
CC	infection.
SQ	Sequence 677 BP; 204 A; 116 C; 142 G; 211 T;
Query Match	16.8%; Score 111; DB 60; Length 677;
Best Local Similarity	62.4%; Pred. No. 1,41e-45;
Matches 265; Conservative 3; Mismatches 157; Indels 0; Gaps 0;	
Db	253 ggaatgataccaaagccgcaaatatcaacagacaacgaagagccatctttgatttta 312
QY	13 GGATGACATTATGCTCAAAATATTTCAAATATGCGATTAAGTAACTGCTGTATGTGTA 72
Db	313 tcacaaggaatgttgtagtgatcatttctgtcttcacagcatttggccaataatcmt 372
QY	73 TGTGGATGGGAGTTTATGATGATCATCTCTGTCACCAATATGCTACCATACGTTT 132
Db	373 gaatcgctgaatgttgtagtgatctgtgcttatttggaaacttgtagtcaaaagtcgc 432
QY	133 AATATGACATTACTCTGAATATGCGATTAGTTAGTTATTAAGAGAGCTAATATGTATCAAAATATTAA 192
Db	433 agaagaatgctaaagttttaaagtgtaaaacgctcgttttagcgagaaaaagtcagtga 492
QY	193 AGAAGCATTTAGAAACGTTTGGTGTGTTAAACGTCGTTTCATTAACATNCAATTGCCAA 252
Db	493 catgattattgttgaatgaattatgcgcacatccacgcagtgaatataaagaacagattgatg 552
QY	253 TCAAGTTATTGTAGATGATTTATGCACACCATCCCAAGAAATATGCTCAATATGACAC 312
Db	553 cgcctcgccaataataatctgcacaagaanaatatattgctgcgtcttcagcgacataaattac 612
QY	313 AGCAGCAAAAGAAATATCCATCAATTAAGAAGTTGTTGCGATTTTCAACCCACACATTTCTC 372
Db	613 acgacaatgtcccttaagtgaatgttcgtcgaagcaactggatttgcagatgaatc 672
QY	373 TAGAACAACAGCATTTTAAATGAATTTGCCAGAAAGTTTAAGTTAAACAGATCTGTATT 432
Db	673 ctat 677
QY	433 CTAT 437
RESULT	6
ID	V27381 standard; DNA; 1267 BP.
AC	V27381;
DT	02-Oct-1998 (first entry)
DE	<i>Streptococcus pneumoniae</i> Sp0070 nucleotide.
KW	<i>Streptococcus pneumoniae</i> ; antigen; vaccine; infection; diagnosis;
OS	detection; pneumonia; otitis media; meningitis; ss.
FS	<i>Streptococcus pneumoniae</i> .
FT	Key location/Qualifiers
FT	CDS 2..1267
FT	/*tag= a
FT	/product= "Sp0070"
FT	/note= "no stop codon given"
FN	WC01891930-A2.











```

PT misincorporation, completion of molecules and screening.
PS Disclosure: P: English.
CC Random point mutations were introduced into the alpha fragment of
CC E. coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 5.5%; Score 36; DB 1; Length 204;
Best Local Similarity 10.1%; Pred. No. 1.24e-05;
Matches 11; Conservative 57; Mismatches 40; Indels 1; Gaps 1;

Db 84 cgmmttth-yrmbvryidynsdaaawcyryrkydcynachddhyvbbv 142
OY 224 CCGCTTTCATGAAACATGCAATTCGAATTCAGTATTGTGATGATTCACACCAT 283
Db 143 yvnhhnncccbnhvchvnhbnnhnmwayrhdarddvhcvcchc 191
OY 284 CCAAGAGAAATTGATGCTACATTTGACACAGCAGCAAGAAATATTCAC 332

RESULT 12
ID Q51746 standard; cDNA; 91 BP.
AC Q51746;
DE 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
DE Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
FN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT ) BECTON DICKINSON CO.
PI Shank DD, Spears PA.
PI WPI: 93-378844/48.
PT New oligo-nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 5.2%; Score 34; DB 9; Length 91;
Best Local Similarity 11.7%; Pred. No. 1.09e-04;
Matches 7; Conservative 40; Mismatches 13; Indels 0; Gaps 0;

Db 12 svhsyvvvvhvshvshvvhvshvsvvvhvshvshvshvsvvsvcaagctcg 71
OY 169 GAAGCAGATGTTACAAATATTAAGAACATTAGAAAGCTTGGTGCTTAACGCTG 228

RESULT 13
ID Q70466 standard; DNA; 114 BP.
AC Q70466;
DE 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR-9 peptide library.
DE TSAR: totally synthetic affinity reagent; synthetic; binding domain;
DE effector domain; concatenated heterofunctional protein; linker;
DE direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.

```

```

FH Key Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN MO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK.
PI WPI: 94-279739/34.
DR P-PSDB: R65152.
DR P-PSDB: R65152.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure: Page 35; 25pp; English.
CC Q70466 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)1(TGC)(NNB)10(TGC)2(NNB)4Z(NNB)8(TGC)(NNB)
CC -9Y. X and Y are flanking restriction sites (X is not the same as Y)
CC that are not specified further. Other generic sequences are shown in
CC Q70466-68. Other specific peptides generated by these generic sequences
CC are shown in R65151-54. TSARs are concatenated heterofunctional proteins
CC or peptides, comprising at least two functional regions - a binding
CC domain with affinity for a ligand and a second effector peptide portion
CC that is chemically or biologically active. They may further comprise a
CC linker peptide between the 2 domains. The oligonucleotides are also
CC designed so that the expressed peptide contains 2 or 4 cysteine residues
CC positioned in, or flanking, the unpredicted or variant residues. These
CC residues confer some degree of conformational rigidity to the peptides.
CC The TSARs or comps. comprising a TSAR binding domain can be used in
CC vivo to deliver a chemically or biologically active moiety, eg. metal
CC ion, radioisotope, peptide, toxin or enzyme, to the specific target or
CC on the cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed actively allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;

Query Match 5.2%; Score 34; DB 12; Length 114;
Best Local Similarity 8.0%; Pred. No. 1.09e-04;
Matches 9; Conservative 30; Mismatches 73; Indels 0; Gaps 0;

Db 3 btgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
CP 350 CTGCACAACTCTTATGTGATATTTCTTTCGTCGCTGTCATTTGACACCAATTT 291
Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
CP 290 CTTTGATGATGCTGATATCATCATCAATTAATCAATTGATTCATTTGACT 239

RESULT 14
ID V44650 standard; DNA; 91 BP.
AC V44650;
DE 06-OCT-1998 (first entry)
DE Mammalian DNA replication origin consensus sequence, uniorsconsensus.
DE DNA replication origin; human; mamma; alphaconsensus; uniorsconsensus;
DE anti-gene; DNA replication inhibitor; shuttle vector construct creation;
DE gene therapy; ss.
OS Mammalia.
PN WO9827200-A2.
PD 25-JUN-1998.
PF 12-DEC-1997; CA0972.
PR 21-MAY-1997; US-047322.
PR 16-DEC-1996; US-033374.
PA (UYMC-) UNIV MCGILL.
PI Cossons NH, Nielsen TO, Price GB, Zannis-Hadjopoulos M;

```



```

CC human telomeres and large centromeric blocks for assembly of human
CC artificial chromosomes and maintained as bacterial plasmids, circular or
CC linear, large or small yeast artificial chromosomes (YACs) or as episomal
CC elements.
50 Sequence 91 BP; 15 A; 1 C; 4 G; 7 T;

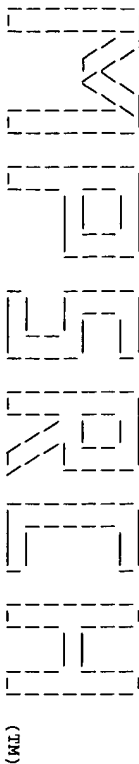
Query Match 4.8%; Score 32; DB 46; Length 91;
Best Local Similarity 12.6%; Pred. No. 9.22e-04;
Matches 11; Conservative 51; Mismatches 24; Indels 1; Gaps 1;

Db 2 wmtvaakravrwwkdkavwggakrkwwkwhassascmkdwkaktwkgswarrrywkgr 61
Cp 618 AANAATGCAATTTTGTATTTTTCGAAATATCACTGCAACCATTAATAAACAACACACATTA 559
Dd
Dd 62 kmwv-twkwawdsdatakwwwkdkwkmv 87
Cp 558 TCAATTTGTTCTAATACATTAATAGAA 532

Search completed: Sat Nov 27 12:35:34 1999
Job time : 168 secs.

```





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_mn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 12:36:19 1999; Maspar time 50.78 seconds

Tabular output not generated. 1124.053 Million cell updates/sec

Title: >US-09-103-287-3  
Description: (1-660) from US09103287.seq  
Perfect Score: 660  
N.A. Sequence: 1 ATTTAAGATTCGATGACA.....GGGTTTATATGTTTATA 660  
Comp: TAAATTTCTAACCTACTGT.....CCCAATATATCAAAATAT

Scoring table: TABLE default  
Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 165359 seqs, 43243793 bases x 2

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: n-issued  
1:5A\_COMB 2:5B\_COMB 3:5C\_COMB 4:PCT9\_COMB 5:backfiles1

Statistics: Mean 7.989; Variance 4.833; scale 1.653

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Prod. No.
C 1	54	8.2	7218	2	US-08-232-	Sequence 14, Applicati	5.22e-19
C 2	29	4.4	215	1	US-08-238-	Sequence 5, Applicatio	1.09e+00
C 3	25	3.8	1200	1	US-08-357-	Sequence 2, Applicatio	3.46e-02
C 4	25	3.8	1200	2	US-08-672-	Sequence 2, Applicatio	3.46e-02
C 5	24	3.6	108	3	US-08-878-	Sequence 2, Applicatio	1.11e-01
C 6	24	3.6	108	2	US-08-357-	Sequence 2, Applicatio	1.11e-01
C 7	24	3.6	965	3	US-08-388-	Sequence 2, Applicatio	1.11e-01
C 8	23	3.5	81	4	PCT-US95-1	Sequence 98, Applicati	3.53e-01
C 9	23	3.5	82	4	PCT-US95-1	Sequence 97, Applicati	3.53e-01
C 10	23	3.5	242	1	US-08-273-	Sequence 1, Applicatio	3.53e-01
C 11	23	3.5	1288	2	US-08-460-	Sequence 9, Applicatio	3.53e-01
C 12	22	3.3	74	4	PCT-US95-1	Sequence 100, Applicat	1.09e+00
C 13	22	3.3	74	4	PCT-US95-1	Sequence 100, Applicat	1.09e+00
C 14	22	3.3	75	4	PCT-US95-1	Sequence 99, Applicati	1.09e+00
C 15	22	3.3	81	4	PCT-US95-1	Sequence 98, Applicati	1.09e+00
C 16	22	3.3	156	1	US-08-545-	Sequence 2, Applicatio	1.09e+00
C 17	22	3.3	156	1	US-08-021-	Sequence 2, Applicatio	1.09e+00
C 18	22	3.3	156	1	US-08-242-	Sequence 2, Applicatio	1.09e+00
C 19	22	3.3	156	1	US-08-084-	Sequence 2, Applicatio	1.09e+00
C 20	22	3.3	156	4	PCT-US94-0	Sequence 2, Applicatio	1.09e+00

21	22	3.3	156	4	PCT-US95-0	Sequence 2, Applicatio	1.09e+00
22	22	3.3	162	1	US-08-242-	Sequence 5, Applicatio	1.09e+00
23	22	3.3	162	4	PCT-US95-0	Sequence 5, Applicatio	1.09e+00
24	22	3.3	162	2	US-08-545-	Sequence 5, Applicatio	1.09e+00
25	22	3.3	774	4	PCT-US91-0	Sequence 1, Applicatio	1.09e+00
26	22	3.3	1011	4	PCT-US96-0	Sequence 119, Applicat	1.09e+00
27	22	3.3	1644	3	US-08-785-	Sequence 4, Applicatio	1.09e+00
28	22	3.3	1662	3	US-08-785-	Sequence 1, Applicatio	1.09e+00
29	22	3.3	3143	4	PCT-US96-0	Sequence 1, Applicatio	1.09e+00
30	22	3.3	3143	2	US-08-485-	Sequence 1, Applicatio	1.09e+00
31	21	3.2	54	3	US-08-452-	Sequence 14, Applicati	3.30e+00
32	21	3.2	65	1	US-08-471-	Sequence 145, Applicat	3.30e+00
33	21	3.2	66	4	PCT-US95-1	Sequence 93, Applicati	3.30e+00
34	21	3.2	68	1	US-08-471-	Sequence 143, Applicat	3.30e+00
35	21	3.2	74	4	PCT-US95-1	Sequence 94, Applicati	3.30e+00
36	21	3.2	75	4	PCT-US95-1	Sequence 99, Applicati	3.30e+00
37	21	3.2	81	4	PCT-US95-1	Sequence 92, Applicati	3.30e+00
38	21	3.2	81	4	PCT-US95-1	Sequence 92, Applicati	3.30e+00
39	21	3.2	215	1	US-08-238-	Sequence 5, Applicatio	3.30e+00
40	21	3.2	516	1	US-08-280-	Sequence 1, Applicatio	3.30e+00
41	21	3.2	965	3	US-08-388-	Sequence 22, Applicati	3.30e+00
42	21	3.2	1498	1	US-08-118-	Sequence 16, Applicati	3.30e+00
43	21	3.2	1498	3	US-08-909-	Sequence 1, Applicatio	3.30e+00
44	21	3.2	1498	3	US-08-909-	Sequence 1, Applicatio	3.30e+00
45	21	3.2	5852	1	US-07-867-	Sequence 2, Applicatio	3.30e+00

## ALIGNMENTS

RESULT 1  
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.  
AC xxxxxx  
DE Sequence 14, Application US/08232463  
CC Sequence 14, Application US/08232463  
CC Patent No. 5670367  
CC GENERAL INFORMATION:  
CC APPLICANT: DORNER, F.  
CC APPLICANT: SCHEIFLINGER, F.  
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
CC NUMBER OF SEQUENCES: 52  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Foley & Lardner  
CC STREET: 1800 Diagonal Road, Suite 500  
CC CITY: Alexandria  
CC STATE: VA  
CC COUNTRY: USA  
CC ZIP: 22313-0299  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/232,463  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/935,313  
CC FILING DATE:  
CC APPLICATION NUMBER: EP 91 114 300.6  
CC FILING DATE: 26-AUG-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: BENT, Stephen A.  
CC REGISTRATION NUMBER: 29, 768  
CC REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (703)836-9300  
CC TELEFAX: (703)863-4109  
CC TELE: 899149  
CC INFORMATION FOR SEQ ID NO: 14:  
CC SEQUENCE CHARACTERISTICS:







```

CC REFERENCE/DOCKET NUMBER: PKR/5544/202253/DJP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-861-3000
CC TELEFAX: 202-822-0944
CC
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1200 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA (genomic)
SQ SEQUENCE 1200 BP; 406 A; 224 C; 230 G; 323 T; 17 OTHER.

Dh 324 GAACAYCAAAAYHHHYAYCAVHAHAHYCCAAACAG 362
    |||:||||:||||:||||:||||:||||:
Cp 95 GATCATTAACCTCACATCCATCACATCATCATCAGCAG 57

RESULT 4
ID US-08-672-514-2 STANDARD; DNA; UNC; 1200 BP.
AC xxxxxx
DT
DE Sequence 2, Application US/08672514
CC Sequence 2, Application US/08672514
CC Patent No. 5686248
CC GENERAL INFORMATION:
CC APPLICANT: BURNIE Mr., James P.
CC APPLICANT: MATTHEWS Ms., Ruth C.
CC TITLE OF INVENTION: FUNGAL STRESS PROTEINS
CC NUMBER OF SEQUENCES: 10
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Cushman, Darby & Cushman
CC STREET: 1100 New York Avenue, N.W.
CC CITY: Washington
CC STATE: DC
CC COUNTRY: USA
CC ZIP: 20005-3918
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/672,514
CC FILING DATE: 28-JUN-1996
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/152,669
CC FILING DATE: 16-NOV-1993
CC APPLICATION NUMBER: US 08/152669
CC FILING DATE: 16-NOV-1993
CC APPLICATION NUMBER: US 663897
CC FILING DATE: 14-MAR-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: KOKULIS, Paul N.
CC REGISTRATION NUMBER: 16773
CC REFERENCE/DOCKET NUMBER: PKR/5544/202253/DJP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-861-3000
CC TELEFAX: 202-822-0944
CC TELEX: 6714627 CUSH
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1200 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA (genomic)

```

```

S0 SEQUENCE 1200 BP; 406 A; 224 C; 230 G; 323 T; 17 OTHER.

Query Match          3.8%; Score 25; DB 2; Length 1200;
Best Local Similarity 43.6%; Pred. No. 3,46e-02;
Matches 17; Conservative 15; Mismatches 7; Indels 0; Gaps 0;

Db      324 GAACATCAAAAYHHHVAAYCCAHNAHAHYHCYCAACAAG 362
        |||||::|::|::|::|::|::|::|::|::|::|::|::|::|
Cp      95 GATCATTAATACTCACATCCACATACACATCAACAGCG 57

RESULT    5
ID US-08-878-106-2 STANDARD; DNA; UNC; 108 BP.
AC xxxxxx

DE Sequence 2, Application US/08878106
CC Sequence 2, Application US/08878106
CC Patent No. 5817466
CC GENERAL INFORMATION:
CC APPLICANT: Hoyer, Lois
CC APPLICANT: Livi, George
CC APPLICANT: Shatzman, Allan
CC TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406-0939
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq Version 1.5
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/878,106
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/357,962
CC FILING DATE: 16-DEC-1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jervis, Herbert H
CC REGISTRATION NUMBER: 31,171
CC REFERENCE/DOCKET NUMBER: P50278
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5019
CC TELEFAX: 610-270-5090
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 108 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: Genomic DNA
CC HYPOTHEtical: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE:
CC ORIGINAL SOURCE:
CC SEQUENCE 108 BP; 20 A; 23 C; 12 G; 10 T; 43 OTHER.

Query Match          3.6%; Score 24; DB 3; Length 108;
Best Local Similarity 31.6%; Pred. No. 1.11e-01;
Matches 18; Conservative 22; Mismatches 17; Indels 0; Gaps 0;

Db      39 RTCNTWGCACHCHACHACHTDACYGSCWCGRNGRNGHACYGAWMCHGTDMT 95
        :||:||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Cp      592 ATCACTCTACCACCTAATATAAACACAGCATTTATCAATTGTTCATATACATTAA 536
```



```

DE      RESULT      6
ID US-08-357-962-2 STANDARD; DNA; UNC; 108 BP.
AC      xxxxxx
DT
Sequence 2, Application US/08357962
CC Sequence 2, Application US/08357962
CC Patent No. 5668263
CC GENERAL INFORMATION:
CC APPLICANT: Hoyer, Lois
CC APPLICANT: Livi, George
CC APPLICANT: Shatzman, Allan
CC TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406-0939
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq Version 1.5
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/357,962
CC FILING DATE: 16-DEC-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jeryis, Herbert H
CC REGISTRATION NUMBER: 31,171
CC REFERENCE/DOCKET NUMBER: P50278
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5019
CC TELEFAX: 610-270-5090
CC TELEX:
CC INFORMATION FOR SEQ ID NO.: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 108 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: Genomic DNA
CC HYPOTHEICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE:
CC ORIGINAL SOURCE:
CC SEQUENCE 108 BP; 20 A; 23 C; 12 G; 10 T; 43 OTHER.
CC
Query Match          3.6%; Score 24; DB 2; Length 108;
Best Local Similarity 31.6%; Pred.No.1,11e-01;
Matches 18; Conservative 22; Mismatches 17; Indels 0; Gaps 0.
Db
39 RTCNTWYGCACHCACHCACHCRTDACYGSCWCWCWRGNGHACYGAYWCHGTDMT 95
   || :|| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Cc 592 ATCACCTCGACCACCAATAATATAAACACAGCATTTATCAATTGTCTAATACATTAA 536

RESULT      7
ID US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.
AC      xxxxxx
DT
Sequence 22, Application US/08388672A
Cc Sequence 22, Application US/08388672A
Cc Patent No. 5795961
Cc GENERAL INFORMATION:
Cc APPLICANT: Wallace, T. Paul
Cc APPLICANT: Harris, William J.
Cc APPLICANT: Carr, Frank J.
Cc
```

```
CC APPLICANT: Old, Lloyd J.  
CC APPLICANT: Melt, Sydney  
CC APPLICANT: Kitamura, Kunio  
CC TITLE OF INVENTION: Recombinant Human Anti-Lewis B  
CC TITLE OF INVENTION: Antibodies  
CC NUMBER OF SEQUENCES: 25  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Felfe and Lynch  
CC STREET: 805 Third Avenue  
CC CITY: New York  
CC STATE: New York  
CC COUNTRY: U.S.A.  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/388,672A  
CC FILING DATE: 14-FEB-1995  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Hanson, No. 5795961man D.  
CC REGISTRATION NUMBER: 30,946  
CC REFERENCE/DOCKET NUMBER: LUD 5409  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 212-688-3884  
CC TELEFAX: 212-838-3884  
CC INFORMATION FOR SEQ ID NO: 22:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 965 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: DNA (genomic)  
CC SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.  
SQ  
  
Query Match 3.6%; Score 24; DB 3; Length 965;  
Best Local Similarity 18.5%; Pred.No. 1,11e-01;  
Matches 20; Conservative 47; Mismatches 40; Indels 1; Gaps 1.  
  
Db 778 TKKURHVVHVSGVSRSTCTASDYTTSTWGVYRGKMGEDYGCGYTNYNGKKRQVTMAD 837  
| : : : : : | : : : : : | : : : : : | : : : : :  
Qy 115 TGGGACCATCAAGTATTAAATGATAGCTGTAATGCATTGATGTATT-TAGAGAAGC 173  
| : : : : : | : : : : : | : : : : : | : : : : :  
Db 838 TSSRSRSVTAADTAIVYYCVGRGSDYDGDGPGTGTVVSSHVKDM 885  
| : : : : : | : : : : : | : : : : : | : : : : :  
Qy 174 TAGATGTTACAATAATTTAAGAACCATTAGAAGCGTTGGTGIGTGA 221  
| : : : : : | : : : : : | : : : : : | : : : : :  
  
RESULT 8  
ID PCT-US95-11934-98 STANDARD; DNA; UNC; 81 BP.  
AC xxxxxx  
DT Sequence 98, Application PC/TUS9511934  
DE Sequence 98, Application PC/TUS9511934  
CC GENERAL INFORMATION:  
CC APPLICANT: Cytozen Corporation  
CC TITLE OF INVENTION: Antigen Binding Peptides (Abtides) From  
CC TITLE OF INVENTION: Peptide Libraries  
CC NUMBER OF SEQUENCES: 103  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Pennie & Edmonds  
CC STREET: 1155 Avenue of the Americas  
CC CITY: New York  
CC STATE: New York  
CC COUNTRY: USA  
CC ZIP: 10036  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS
```



```
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/11934
CC FILING DATE: 20-SEP-1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Mistrock, S. Leslie
CC REGISTRATION NUMBER: 18,872
CC REFERENCE/DOCKET NUMBER: 1101-196-228
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 790-9090
CC TELEFAX: (212) 869-9741/8864
CC TELEX: 66141 PENNIE
CC INFORMATION FOR SEQ ID NO: 98:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 81 base pairs
CC MOLECULE TYPE: DNA (genomic)
CC SEQUENCE: 81 BP; 6 A; 6 C; 4 G; 5 T; 60 OTHER.
SQ
Query Match 3.5%; Score 23; DB 4; Length 81;
Best Local Similarity 10.1%; Pred. No. 3.53e-01;
Matches 7; Conservative 19; Mismatches 43; Indels 0; Gaps 0.
Db 5 TGCTGAGANNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNV 64
||| |||::: : : : : : : : : : : : : : : : : : : : : : : :
Cc 111 TGTGAGACAGAGAAGTGATCATTAACAACACCAATCACATACATCAGACGACTACT 52
Dc 65 NNNNNNNNA 73
Cc 51 TTATCCGTA 43
Pct-us95-11934-97 STANDARD: DNA: UNC: 82 BP.
RESULT 9
ID PCT-US95-11934-97 STANDARD: DNA: UNC: 82 BP.
DI xxxxxx
DE Sequence 97, Application PC/TUS9511934
CC Sequence 97, Application PC/TUS9511934
CC GENERAL INFORMATION:
Cc APPLICANT: Cytogen Corporation
Cc TITLE OF INVENTION: Antigen Binding Peptides (AbTides) from
Cc TITLE OF INVENTION: Peptide Libraries
Cc NUMBER OF SEQUENCES: 103
Cc CORRESPONDENCE ADDRESS:
Cc ADDRESSEE: Pennie & Edmonds
Cc STREET: 1155 Avenue of the Americas
Cc CITY: New York
Cc STATE: New York
Cc COUNTRY: USA
Cc ZIP: 10036
Cc COMPUTER READABLE FORM:
Cc MEDIUM TYPE: Floppy disk
Cc COMPUTER: IBM PC compatible
Cc OPERATING SYSTEM: PC-DOS/MS-DOS
Cc SOFTWARE: Patentin Release #1.0, Version #1.30
Cc CURRENT APPLICATION DATA:
Cc APPLICATION NUMBER: PCT/US95/11934
Cc FILING DATE: 20-SEP-1995
Cc CLASSIFICATION:
Cc ATTORNEY/AGENT INFORMATION:
Cc NAME: Mistrock, S. Leslie
Cc REGISTRATION NUMBER: 18,872
Cc REFERENCE/DOCKET NUMBER: 1101-196-228
Cc TELECOMMUNICATION INFORMATION:
Cc TELEPHONE: (212) 790-9090
Cc TELEFAX: (212) 869-9741/8864
Cc TELEX: 66141 PENNIE
Cc INFORMATION FOR SEQ ID NO: 97:
Cc SEQUENCE CHARACTERISTICS:
Cc LENGTH: 82 base pairs
```

[illegible]







CC	MOLECULE TYPE:	DNA (genomic)
50	SEQUENCE 75 BP; 1 A; 1 C; 7 G; 5 T; 61 OTHER	

Search completed: Sat Nov 27 12:37:13 1999  
Job time : 54 secs.



**THIS PAGE BLANK (USPTO)**



\*\*\*\*\*  
 Release 3.1a John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd  
 \*\*\*\*\*  
 (TM)

\*\*\*\*\*  
 Release 3.1a John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd  
 \*\*\*\*\*

Merch\_ntp n.a. - n.a. Smith-Waterman search, using a protein database  
 which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Nov 24 02:04:43 1999; Maspar time 108.96 Seconds  
 1456.244 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-3  
 Description: (1-660) from US09103287.seq  
 Perfect score: 3300

N.A. Sequence: 1 ATTTAAGATTCGGATGACA.....GGTTTATATATGTTATTA 660  
 Comp: TAAATTTCACGCTACTGT.....CGCAAAATTATACAAATTT

Scoring table: TABLE bkttranslated2  
 Gap 30

Mmatch STD : Dbase 0; Query 0

Searched: 122810 seqs, 120205779 bases x 2

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: p1r60

1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 73.220; Variance 175.053; scale 0.418

Prod. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1323	40.1	432	2	C69662	UDP-N-acetylmuramate-2,86e-16	
2	313	9.5	454	2	B70418	UDP-N-acetylmuramate-2,92e-21	
3	306	9.3	491	1	CEPCAM	UDP-N-acetylmuramate-2,82e-20	
4	296	9.0	475	2	E64185	UDP-N-acetylmuramate-6,96e-19	
5	294	8.9	468	2	H70201	UDP-N-acetylmuramate-1,32e-18	
6	284	8.6	505	2	S76722	UDP-N-acetylmuramoyla-3,13e-17	
7	269	8.2	471	2	A71475	Probable muramate-Ala 3,39e-15	
8	252	7.6	481	2	C71338	Probable UDP-N-acetyl-6,15e-13	
9	213	6.5	495	2	C71679	UDP-n-acetylmuramate-5,51e-08	
10	198	6.0	449	2	B71917	UDP-n-acetylmuramate-3,48e-06	
11	180	5.5	449	2	G64597	UDP-n-acetylmuramate-4,06e-04	
12	172	5.2	494	2	D70579	Probable murC protein-3,09e-03	
13	167	5.1	251	1	HBSHBC	hemoglobin beta (CNA)-1,06e-02	
14	168	5.1	516	2	S34525	hypothetical protein-8,32e-03	
15	163	4.9	398	1	R3BVM1	ribosomal protein var-2,81e-02	
16	157	4.8	99	2	S44632	f22b7.3 protein - Cae-1,17e-01	
17	157	4.8	703	2	T64226	cell division control-1,17e-01	
18	157	4.8	758	2	S45825	cell division control-1,17e-01	
19	159	4.8	1182	2	G71607	probable integral mem-7,30e-02	

RESULT ENTRY TITLE	1	156	155	154	153	152	151	150	149	148	147	146	145	144	143
ORGANISM	1	195	195	195	195	195	195	195	195	195	195	195	195	195	195
DATE	05-Dec-1997	05-Dec-1997	05-Dec-1997	05-Dec-1997	05-Dec-1997	05-Dec-1997	05-Dec-1997	05-Dec-1997	05-Dec-1997	05-Dec-1997	05-Dec-1997	05-Dec-1997	05-Dec-1997	05-Dec-1997	05-Dec-1997
ACCESSIONS	C69662	C69662	C69662	C69662	C69662	C69662	C69662	C69662	C69662	C69662	C69662	C69662	C69662	C69662	C69662
REFERENCE	#authors	#authors	#authors	#authors	#authors	#authors	#authors	#authors	#authors	#authors	#authors	#authors	#authors	#authors	#authors
ALIGNMENTS	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC -
spore coat protein co	1.48e-01	1.48e-01	1.48e-01	1.48e-01	1.48e-01	1.48e-01	1.48e-01	1.48e-01	1.48e-01	1.48e-01	1.48e-01	1.48e-01	1.48e-01	1.48e-01	1.48e-01
DNA-directed RNA poly	1.87e-01	1.87e-01	1.87e-01	1.87e-01	1.87e-01	1.87e-01	1.87e-01	1.87e-01	1.87e-01	1.87e-01	1.87e-01	1.87e-01	1.87e-01	1.87e-01	1.87e-01
NADH dehydrogenase (u	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01
immunogenic protein p	2.96e-01	2.96e-01	2.96e-01	2.96e-01	2.96e-01	2.96e-01	2.96e-01	2.96e-01	2.96e-01	2.96e-01	2.96e-01	2.96e-01	2.96e-01	2.96e-01	2.96e-01
conserved hypothetical	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01	4.68e-01
hypothetical protein	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01
hemoglobin beta (CNA)	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01
conserved hypothetical	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00
ubiquinol--cytochrome	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01
thymidine kinase (tdk	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01
ribosomal protein VAR	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01
asparagine-rich prote	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01	7.35e-01
hypothetical 58.0K pr	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01
G-box-binding factor	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01	5.87e-01
cell division control	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00
hypothetical protein	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01	9.20e-01
DNA-directed RNA poly	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00	1.15e+00
ribosomal protein var	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00
GTP-binding regulator	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00
mRNA maturase b12 - y	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00
hypothetical protein	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00	1.44e+00
coit intron protein 2	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00
ATP-dependent protein	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00	2.23e+00
hypothetical protein	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00	1.79e+00
ubiquinol--cytochrome	2.77e+00	2.77e+00	2.77e+00	2.77e+00	2.77e+00	2.77e+00	2.77e+00	2.77e+00	2.77e+00	2.77e+00	2.77e+00	2.77e+00	2.77e+00	2.77e+00	2.77e+00

RESULT	1	
ENTRY	C69662	#type complete
TITLE	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) murC - Bacillus subtilis	
ORGANISM	#formal_name Bacillus subtilis	
DATE	05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 17-Mar-1999	
ACCESSIONS	C69662; S71002	
REFERENCE	A69580	
#authors	Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertolo, M.G.; Bessières, P.; Bolojin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devane, K.M.; Duesterhoef, A.; Enrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erlington, J. Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G. Ghispi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Kleier-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, S.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Punnett, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Rocha, B.; Rose, M.; Sadale, Y.; Sato, T.; Scallion, E.; Schleich, S.; Schroeter, R.; Scottone, F.; Sekiyuchi, J.; Sekowska, A.; Serot, S.J.; Serrito, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takanaru, K.; Takeuchi, M.; Tamashiki, A.; Tanaka, T.; Terstrop, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambolt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.	
ALIGNMENTS		



[illegible]

Db	1040	GAACACGACATTCTTTTNCAGARTTTGCGNAGRMSNYTMSGNGCAGATYCTNTAY	1099
Dc	375	GAACACGACATTCTTTTNAATGAATTCGCGAAGTTTGAAGTAAAGACGATCGTATCT	434
Qc		T Q A F L N E F A E S L S K A D R V F L	
Db	1100	TNTGCAATHTTTCGMSNCGNNGNARAYCGCNGCAATYTTACNAETHGNCATYTTNC	1159
Dc	435	TATGTGAATTTTGGATGACATTCAGAAATTAAGTGGCGCATTAACATCAATTTAA	494
Qc		C E I F G S I R E N T G A L T I O D L I	
Db	1160	ARGNAARATHCATNAAYCCNARTTNATHGARGARGAYGATACMWSGNTYNTNARBCNC	1219
Dc	495	TTGATTAATTAATGAAGTCGATCGTTAATATGAAAGATTCATTATGATTTAGCAACAAT	554
Qc		D K I E G A S L I N E D S I N V L E Q F	
Db	1220	AYGATAAAGCNGTNTNTNATHHTTATGGCNGCNGGNGATTCACAAT	1267
Dc	555	TTGATAATGCGTGTGTTTATTTGGTGCAGGCGATATTCACAAAAT	602
Qc		D N A V V L F M G A G D I Q K	
RESULT	2		
ENTRY	B70418	#type complete	
TITLE	UDP-N-acetylilmurate-alanine ligase - Aquifex aeolicus		
ORGANISM	#formal name Aquifex aeolicus		
DATE	08-May-1998	#sequence_revision 08-May-1998	#text_change 21-Aug-1998
ACCESSIONS	B70418		
REFERENCE	A70300		
authors	Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;		
	Kellom, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.;		
	Kellom, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,		
	J.M.; Olson, G.J.; Swanson, R.V.		
#journal	Nature (1998) 392:353-358		
#file	The complete genome of the hyperthermophilic bacterium		
	Aquifex aeolicus.		
#cross-references	MUID:98196666		
#accession	B70418		
#status	preliminary; nucleic acid sequence not shown;		
	translation not shown		
#molecule_type	DNA		
##residues	1-454	##label AOF	
##cross-references	GB:AE000736; NID:g2983763; PID:g2983764; GB:AE000655		
##experimental_source	strain VFS		
GENETICS			
#gene	murC		
CLASSIFICATION	#superfamily UDP-N-acetylilmurate- alanine ligase		
SUMMARY	#length 454 #molecular-weight 50893 #checksum 2346		
Query Match	9.5%; Score 313; DB 21; Length 454;		
Best Local Similarity	39.4%; Pred. No. 2,92e-21;		
Matches 117; Conservative	54; Mismatches 121; Indels 5; Gaps 5;		
Db	H N V Y N A L A A T G V A L E L G V S F		
Dc	831 HCAYAAAGTNTAAVAGCANTGNCNGCAGCNG-CNGTNGCNTNGARTYNGCNGTMSNT	889	
Qc	121 CCATPCAGTTTAAATGATAGCTGTAATTCGCATTAATTTAGAGAAGCTAGTGT	180	
	H T V L N A L A V I A L V I * R S * M L		
Db	E V I K K S L T N E E F R N A E R R L E L K		
Dc	890 TYGARGCTAATAAARASNTLNBARARTTYMGNALYCGNARGRMGMYTNGARTNA	949	
Qc	181 TACAAATTAAGAAAG-CATTTAAGAACGTTTGCTGCTGTTAAACGCTTC-AATGA	238	
	G Y I L L K K X L E T F G G V K R R F X * N		
Db	G Y Y K N S P V Y D D Y G H H P T E I K		
Dc	950 ARGNTATATAAARAAVSMSCGNTNTAYGATGATYTAGCNCAYCATCCACNAGARTHA	1009	
Qc	239 ACTACAAATTCGAATTCAGATTAT-TGTA-CATGCTTATTCGACACCATCCACGAAGAAATTA	296	



Dc		A V I N S L R D M Y P D K N L L V F Q    ::: 1010 ARGCGNTHAAWNSITNMGNAVYGTAYCCNGVAARAAVTNYTNGINTTYC 1069
Oy	297	GTCGTACATGCAGCACGAAGAATTCCCATTAACAAGACTGGTGCACTTTC 356 A T I D T A R K K Y P H K E V V A V F Q
Pt		
Dd		P H R Y S R T Y Y L F E D F V K V L
Dc	1070	ARCNCAYMNTAVWSMNMCNATYAITYTTTTYGARGYTYYGYMAARGNYTNA 1126     ::: OY AACCAACACTTCTTAGAACACACCATTTTTTAATGATAATTCAGAAAGTTAA 413 P H T F S R T Q A F L N E F A E S L
Rresult	3	CCECM #type complete ENTRY UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) - TITLE Escherichia coli
Alternate_names		UDP-N-acetylmutamoyl-L-alanine synthetase ORGANISM #formal_name Escherichia coli DATE 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 02-Jul-1998
Accessions	J00545; S65354; C64731	
REFERENCE	J00544	Ikeba, M.; Wachi, M.; Jung, H.K.; Ishino, F.; Matsubashi, M.
#authors	Nucleic Acids Res. (1990) 18:4014	
#journal	Nucleotide sequence involving murg and murg in the mre gene cluster region of Escherichia coll.	
#title		
#cross-references	MUID:90326550	
#accession	J00545	
molecule_type	DNA	
#residues	1-491 ##label IKE	
REFERENCE	S65354	X52644; NID:g42053; PID:g42056
#cross-references	EMBL:	
#authors	Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A. Fujita, N.; Isoho, K.; Mizouchi, K.; Nakata, A.	
#submission	submitted to the EMBL Data Library, December 1992	
#accession	S40601	
molecule_type	DNA	
#residues	1-491 ##label YUR	
REFERENCE	S65354	D10483; NID:g216434; PID:d1001828; PID:g216505
#cross-references	EMBL:	
#authors	Liger, D.; Masson, A.; Blanot, D.; van Heijenoort, J.; Parquet, C.	
#journal	Eur. J. Biochem. (1995) 230:80-87	
#title	Over-production, purification and properties of the uridine-diphosphate-N-acetylmurate:L-alanine ligase from Escherichla coll.	
#cross-references	MUID:95324553	
#accession	S65354	
#status	preliminary	
molecule_type	protein	
#residues	1-14 ##label LG	
REFERENCE	A64720	
#authors	Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glaeser, J.D.; Rodde, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao Y.	
#journal	Science (1997) 277:1453-1462	
#title	The complete genome sequence of Escherichia coli K-12.	
#cross-references	MUID:9742617	
#accession	C64731	
#status	preliminary; nucleic acid sequence not shown; translation not shown	
molecule_type	DNA	
#residues	1-491 ##label BLAT	
#cross-references	GB:AEO00118; GB:U00096; NID:g1786262; PID:g1786279;	
GENETICS	UMCP:D0091	
#experimental_source	strain K-12, substrain MG1655	
gene	murg	

```

#map position 2 min
FUNCTION
#description one of the ligases responsible for the synthesis of
UDN-acetylmuramyl pentapeptide, an intermediate in
cell-wall biosynthesis
#pathway peptidoglycan biosynthesis
CLASSIFICATION superfamily UDP-N-acetylmuramate--alanine ligase
KEYWORDS ATP; cell division; cell wall; ligase; P-loop; peptidoglycan
biosynthesis
FEATURE
126-131
SUMMARY #region nucleotide-binding motif A (P-loop)
#length 491 #molecular-weight 53626 #checksum 8296
Query Match 9.3%; Score 306; DB 1; Length 491;
Best Local Similarity 43.0%; Pred. No. 2.82e-70;
Matches 80; Conservative 27; Mismatches 79; Indels 0; Gaps 0;
Dc V D D Y G H H P T E V D A T I K A A R A
Dt 1043 TNGTNGAGVATYAGNCCATYCCNCGNARGTNGVYGNACNATHAARGCNGMNG 1102
Qy 261 TTGTGATGATATGACACACCATCAAGAAATTAIGCTACAAATGACACAGCA 320
Qt V D D Y A H H P R E I S A T I D I T A R K
Dc G W P D K N L V M L F Q P H R F T A R T R
Dt 1103 CNGGNTGGCCNGAFAARAAYTNGTNTGYNTTYACRCNCAYMGTTYACMNGNACNN 1162
Qy 321 AGAATATCCACATTAAGAAAGTTTTCAGATTTTCAACCAACCACTTCTCTAGAACA 380
Qt K Y P H K E V V A V F Q P H T F S R T Q
Dc D L Y D D F A N V L T Q V D T L L M L E
Dt 1163 GNGAYTATGAGVATYTCGNAYSTNTNACNARGNCAVACVNTYATGATGNG 1222
Qy 381 AAGCATTTTAAATGATTTCCAAAGATTTAAGTAAAGCAGATCGTGATTTCTTAGTG 1440
Qt A F L N E F A E S L S K A D R V F L C E
V
Dc 1223 ARGINT 1228
Dt 1 1 1 1
Qy 441 AAATTT 446
Qt I

```

```

RESULT      4
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MD1:95350630
#accession
#status
#molecule_type DNA
#residues 1-475 #label TIGR
#cross-references GB:IL4203; NID:91574694; PID:91574655;
TIGR:H11139
#description one of the ligases responsible for the synthesis of
UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) -
Haemophilus influenzae (strain Rd Kx20)
#formal_name Haemophilus influenzae
18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
24-Oct-1998
E64185 #type complete
UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8) -
Haemophilus influenzae (strain Rd Kx20)
#formal_name Haemophilus influenzae
18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
24-Oct-1998
E64185
A64000
Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
Kirkness, E.F.; Kerlavage, A.R.; Bult, C.U.; Tomb, J.F.;
Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
Fitzhugh, W.; Field, C.; Gocayne, J.D.; Scott, J.;
Shirley, R.; Liu, L.; Glodek, A.; Kelley, J.M.; Weidman
J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
M.D.; Utterback, T.R.; Hanne, M.C.; Nguyen, D.T.; Sadek,
D.N.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Furman
J.L.; Georgagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
Science (1995) 269:466-512
Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd

```







Ot	D D I Y A Q I F Q I T D K G T A V D V Y
Dd	E R G V C L G T M T V T L P G D H N I S
Dc	921 GGAMRGNGCMTNTGTYTNGNACMNTGCNCGTNNACNTTCNGSGNAVCAAYATHWS 980
OY	73 TGTCGAGTGCGAGTTGATTATCATCTCCGTCGTCGCACAAATATGGTAGCACATAAGCTTT 132
Ot	V D G E F Y D H V G F R L S P Q Y G D H T V L
Dd	N A L A V A V A V A V G L S L O L D F P V I A
Dc	981 NAATGCTNTGCNCNCCTNCCTNGCNMGNTYTNTNGSNTYNGATYITCCNGTNATHGC 1040
OY	133 AAATGACTTGCGCTTAATTCGATTTGGTATTATTAGAAGCTAGATGTGTACAATAATGTA 192
Ot	N A L A V I A I S Y L E K L D V T N I K
Dd	Q A I A S F N G G A K R R F E C K G Y C N
Dc	1041 NCARGCANATGCKNSNTTAAHYGNGCNMAKRMNMONTTYTGARTGAARAGCTATYGTAA 1100
OY	193 AGAAGACTTGAAGAACCTGTGGTGCTTTAAAGCTCGTTTC -AATGAACCTCAATATGCA 251
Ot	E A L E E T F G G V K R R F x * N Y N C K
Dd	G I T F I D D Y A H H P S E L L A T L A
Dc	1101 YGNANATACNTTATTHAGYAATATXGNCNCAIYACCNCBMSNGARTYNTYNGCNACHTYTGCC 1160
OY	252 ATCAAGTT-ATTC-TAATATGATTATTCACACCTCCACAGAAGAAATTAAGTGTACATTTGA 309
Ot	S S X L x D D Y A H H P R E I S A T I D
Dd	A A K Q K V T H
Dc	1161 NGCGNACBARABGTNNCNAV 1185
OY	310 CACAGCACGACGAAGAAATATCCACAT 334
Ot	T A R K K K Y P H

RESULT	7	
ENTRY		
TITLE	A71475	#type complete
ORGANISM		probable muramate-Ala ligase and D-Ala-D-Ala ligase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
DATE		#formal_name Chlamydia trachomatis 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 21-Nov-1998
ACCESSIONS	A71475	
REFERENCE	A71570	
authors		Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
#journal		Science (1998) 282:754-759
#title		Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.
#cross-references	MUID:99000809	
#accession	A71475	
##status		preliminary
##molecule_type		DNA
##residues		1-803 #label ANR
##cross-references	GB:AE001348; GB:AE001273; NID:g3329216; PID:g3329222	
##experimental_source		serotype D, strain UW-3/Cx
GENETICS		
gene	murC/dalA	
SUMMARY	#length 803	#molecular-weight 89229 #checksum 235
Query Match	8.2%	Score 269; DB 2; Length 803;
Best Local Similarity	33.9%	Pred. No. 3,39e-15;
Matches 123;	Conservative 64;	Mismatches 173; Indels 3; Gaps 3.
Db	Y A D I E B V Q L V G M H N V L N A A A A	
Dc	767 TTTAATGCAATATGACATATGCTGACCATATGTTTAAATGCAATTA	826
Qc	TTTATGATCATTCTCTGCTCCACATATGGTACCATACGTTTAAATGCAATTA	146
Qd	Y D H F L S P Q Y G D H T V L N A L A V	
Db	M G I A L S L G I D E G A I R N A F R G	
Dc	827 CNAATGGNAATHCNCNTMNSNTTNGNAATHGAYGARGSGCNAATMGNAATCNTTYGNG	886
Qc	TTTATGCAATATGTTTATGAGAACTAGATGTTTACAAATATTAAGAAACATTTAGAA	206

```

Qc I A I S Y L E K L D V T N I K E A L E T
Db F S G V Q R R L O R K N S S E T F L F L
Dt 887 GNTTYSNGNGTCARGMNMGNTNCARMGNARAAYWSMWSNGARACNTTYTNTTY 946
   || : | | | : | : | : | : | : | : | : | : | : | : | : | : | :
Oy CGTTGGGTCTTAACGTCGTT -CAATCAACTCAATTCGAATCAAGT -ATTGT 264
Qt F G G V K H R X Q * N Y C K S X L X

E D Y A H H P S E I S C T L R A V R T A
Dc 947 TNGARGATTAYGCAYCAYCCNSNBARITWSNTYACNYTNWNGCNTNGNACNG 1006
   || : | | | : | : | : | : | : | : | : | : | : | : | : | : | :
Oy 265 -AGATGATTATGACACCATTCCAGAAATAATAGTCTAATGACAACAGCAGAAGA 323
Qt D D Y A H H P R E I S A T I D T A R K K

V G Q R R I L A I Y Q P H M R F S R L R E
Db 1007 CNGTNGCACRBMGMNGNTHYTNCNATHIYCARCCNCAHMGNTYSSNRLTNGNG 1066
Dc
Oy 324 AATATPCCAATPAAGAGTAGTGTTCCAGTATTTCAACACACACACTTCTCRAGAACACAG 383
Qt Y P H K E V V A V F Q P H T F S R T Q A

C I D S F P S A F K D A D E V L L T F E V
Db 1067 ARTGATIGAWSTNTTYCCNMSNCGNCTTYARRGVAGCNGAVGANGTYNTTYNACNARG 1126
Dt
Oy 384 CATTTTTAAAGAAITTGCCAGAAAGTTTAAGTAAAGACGATCGTGATTCGTATGATGTA 443
Qt F L N E F A E S L S K A D R V F L C E I

TNT TNT 1129
Db
Dt
Oy 444 TTT 446

```

```

RESULT      8
ENTRY       C71338          #type complete
TITLE       probable UDP-N-acetylglucosamine-6-phosphate 1-glucose 6-phosphate transferase (murC) - syphilis spirochete
ORGANISM    #formal_name Treponema pallidum subsp. pallidum #common_name syphilis spirochete
DATE        24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 17-Mar-1999
ACCESSIONS  C71338
REFERENCE   A71250
AUTHORS     Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Swim, M.; Hickey, E.K.; Clayton R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod, M.P.; Salzberg, S.; Peterson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uffendack, T.; McDonald L.; Artach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.; Garland, S.; Hatch, B.; Horst, C.; Roberts, K.; Wattey, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

#journal    Science (1998) 281:375-388
#title      Complete genome sequence of Treponema pallidum, the syphilis spirochete.
#cross-references MIMD:98332770
#accession  C71338
#status     preliminary; nucleic acid sequence not shown; translation not shown

##molecule_type DNA
##residues    1-481 ##label COL
##cross-references GB:AEO01213; GB:AEO00520; NID:g3322606; PID:g3322616
##experimental_source strain Nichols

TP0341
SUMMARY     TP0341 #length 481 #molecular-weight 53762 #checksum 6883

Query Match      7.6%; Score 252; DB ?; Length 481;
Best Local Similarity 37.4%; Pred. No. 6.15e-13;
Matches          76; Conservative 36; Mismatches 91; Indels 0; Gaps 0;

```



DE	1057	GCMACNTYNGCAGCGAARAA	1079
OY	299	GCTACATTGACACGCAGAA	321
Qc	A T I D T A R		
RESULT	10		
ENTRY	B71917	#type complete	
TITLE	udp-N-acetylmutarinate- alanine ligase - Helicobacter pylori (strain J99)		
ORGANISM	#formal_name Helicobacter pylori		
DATE	12-Feb-1999	#sequence_revision 12-Feb-1999	#text_change 12-Feb-1999
ACCESSIONS	B71917		
REFERENCE	A71800		
authors	Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge, B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.; Uria-Nickelsen, W.; Mills, D.M.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Trust, T.J.		
#journal	Nature (1999) 397:176-180		
#file	Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori.		
#cross-references	MUID:99120557		
#accession	B71917		
#status	Preliminary		
#molecule_type	DNA		
#residues	1-449 ##label ARN		
#cross-references	GB:AE001489; GB:AE001439; NID:g4155102; PID:g4155106		
GENETICS	#experimental_source strain J99		
#gene	muc		
SUMMARY	#length 449 #molecular-weight 50906 #checksum 3196		
Query Match	6.0%; Score 198; DB 2; Length 449;		
Best Local Similarity	39.2%; Pred. No. 3,48e-06;		
Matches	65; Conservative 34; Mismatches 65; Indels 2; Gaps 2;		
DB	S A L D E L N L E E I R N N L L N F K G		
DC	842 TNMSGCYTGTGAAGARTNAAYTTNARGARATHMGNAAVYTYTNAATTTAAAG 901		
OY	156 TTGGTTATTGAGAAGCTAGACTTTCAATAATTAAGAAGCATTAAGAACGTTGGTG 215		
QC	S Y L E K L D V T N I K E A L E T F G G		
DB	I K K R F D I L O K N N L I L I L I D D Y		
DC	902 GNAHHAARAAMGTTYGAY-ATHYTTCARAARAAYTAATYNTHYNATHAGAYATAY 960		
OY	216 GTGTAAACGCCGCTTAATGAACAATTCATTCGA-AATCAATTATGTAGATGTAT 274		
QC	V K R R F F N K L Q L X N Q V I V D D Y		
DB	A H H P T E I G A T L K S A R		
DC	961 GCNCAYCAIYCNCACNGARATHGNGCNACAYTTNAPRSNCGNMGNA 1006		
OY	275 GCACACCATCCAGAGAATTAGTGCATTCAGCACACACAGAA 320		
QC	A H H P R E I S A T I D T A R		
RESULT	11		
ENTRY	G64597	#type complete	
TITLE	UDP-N-acetylmuramate-alanine ligase - Helicobacter pylori (strain 26655)		
ORGANISM	#formal_name Helicobacter pylori		
DATE	09-Aug-1997	#sequence_revision 09-Aug-1997	#text_change 10-Oct-1997
ACCESSIONS	G64597		
REFERENCE	A64520		
authors	Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klein, H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Gloeck, A.;		

DE	1057	GCGACATTTGTCAGCAGCAAA	1079
OY	299	GCTACATTGACACGACGAAA	321
QC		A T I D T A R	
RESULT	10		
ENTRY	B71917	#type complete	
TITLE	udp-N-acetylmutarinate- alanine ligase - Helicobacter pylori (strain J99)		
ORGANISM	#formal_name Helicobacter pylori		
VARIETY	strain J99		
DATE	12-Feb-1999	#sequence_revision 12-Feb-1999	#text_change 12-Feb-1999
ACCESSIONS	B71917		
REFERENCE	A71800		
AUTHORS	Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonghe, B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.; Uria-Nickelsen, W.; Mills, D.M.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Trust, T.J.		
JOURNAL	Nature (1999) 397:176-180		
FILE	Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori.		
CROSS-REFERENCES	MUID:99120557		
ACCESION	B71917		
STATUS	Preliminary		
MOLECULE TYPE	DNA		
RESIDUES	1-449 ##label ARN		
CROSS-REFERENCES	GB:AE001489; GB:AE001439; NID:g4155102; PID:g4155106		
EXPERIMENTAL SOURCE	Strain J99		
GENETICS			
Gene	muc		
SUMMARY	length 449 #molecular-weight 50906 #checksum 3196		
Query Match	6.0%; Score 198; DB 2; Length 449;		
Best Local Similarity	39.2%; Pred. No. 3,48e-06;		
Matches	65; Conservative 34; Mismatches 65; Indels 2; Gaps 2;		
DB	S A L D E L N L E E I R N N L L N F K G		
DC	842 TNMSGCYTGTGAAGARTTAATTAARARATHMGNAAAYTYNTAATTTAAAG 901		
OY	156 TTGGTTATTGAGAAGCTAGATTTCAATAATTAAGAAGCATTAAGAAGCTTGTTG 215		
QC	S Y L E K L D V T N I K E A L E T F G G		
DB	I K K R F D I L O K N N L I L I D D Y		
DC	902 GNAHHAARAAMGTTTGAY-ATHYTNCARAARAAYTNTHYNATHAGAYATAY 960		
OY	216 GTGTAAACGCCTTTAATGAACAATTCATTCGA-AATCAATTATGTAGATGTAT 274		
QC	V K R R F F N K L Q L X N Q V I V D D Y		
DB	A H H P T E I G A T L K S A R		
DC	961 GCNCAYCAIYCNCACNGARATHGNGCNACAYTNAPRSNCGNMGNA 1006		
OY	275 GCACACCATCCAGAGAAATTAGTCATTCAGCACACACAGCAA 320		
QC	A H H P R E I S A T I D T A R		
RESULT	11		
ENTRY	G64597	#type complete	
TITLE	UDP-N-acetylmurinate-alanine ligase - Helicobacter pylori (strain 26655)		
ORGANISM	#formal_name Helicobacter pylori		
DATE	09-Aug-1997	#sequence_revision 09-Aug-1997	#text_change 10-Oct-1997
ACCESSIONS	G64597		
REFERENCE	A64520		
AUTHORS	Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klein, H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Gloeck, A.;		



Matches	0:	Conservative	10:	Mismatches	15:	Indels	0:	Gaps	0:
Dt	1048	GAGGATYAYCCNACAYCCNACGARRARHACGACNACNYTNGCGCGCNGGNA	1102						
QY	266	GATGATATATCACACATCCACAGAAATATAGCTACTGATATGACACAGACGAA	320						
Qt		D D Y A H H P R E I S A T I D T A R							
RESULT	13								
ENTRY		HBSHBC	#type complete						
TITLE		hemoglobin beta C(VA) chain - aoudad							
ORGANISM		#formal_name Ammotragus lervia #common_name aoudad, Barbary sheep							
DATE		24-Apr-1984 #sequence_revision 30-Sep-1988 #text_change 10-Jul-1998							
ACCESSIONS		A90239; A02397							
REFERENCE		A90239							
#authors		Huisman, T.H.J.; Dasher, G.A.; Moretz Jr., W.H.; Dozy, A.M.; Wilson, J.B.; Vliet, G.V.							
#journal		Biochem. J. (1968) 107:745-751							
#title		Studies of haemoglobin types in Barbary sheep (Ammotragus lervia).							
#contents		Composition of tryptic peptides							
#accession		A90239							
#molecule_type		protein							
#residues		1-141 #label HUI							
REFERENCE		A94566							
#authors		Huisman, T.H.J.							
#submission		submitted to the Atlas, January 1969							
#contents		annotation; revision to residue 103							
COMMENT		This type of beta C chain is found in nonanemic Barbary sheep, whereas the other type of beta C chain is found in anemic Barbary sheep.							
CLASSIFICATION		#superfamily globin: globin homology							
FEATURES		blood; chromoprotein; erythrocyte; heme; iron; oxygen carrier							
FEATURE									
1-141									
58		#domain globin homology #label GDB\							
		#binding_site oxygen (His) (distal axial ligand) #status predicted\							
87		#binding_site heme iron (His) (proximal axial ligand) #status predicted							
		#length 141 #molecular_weight 15633 #checksum 8915							
SUMMARY									
		Insufficient memory To Reconstruct alignment.							
RESULT	15								
ENTRY		R3BYM1	#type complete						
TITLE		ribosomal protein vari, mitochondrial - yeast (Saccharomyces cerevisiae) mitochondrion (SGC2)							
ALTERNATE_NAMES		protein 00140							
ORGANISM		#formal_name mitochondrion Saccharomyces cerevisiae							
DATE		18-Apr-1984 #sequence_revision 29-Jan-1999 #text_change 01-Feb-1999							
ACCESSIONS		S78670; S78671; A02751							
REFERENCE		S78634							
#authors		Foury, F.; Roganti, T.; Lecrenier, N.; Purnelle, B.							
#submission		submitted to the Protein Sequence Database, December 1998							
#accession		S78670							
#molecule_type		DNA							
#residues		1-398 #label FOU1							
#cross-references		EMBL:AJ011856; MIPS:00140							
REFERENCE		##experimental_source strain FY1679, isogenic derivative of strain S288c							
#authors		Z13743							
#journal		Foury, F.; Roganti, T.; Lecrenier, N.; Purnelle, B.							
#title		FEBS Lett. (1998) 440:325-331							
		The complete sequence of the mitochondrial genome of Saccharomyces cerevisiae.							
#accession		S78671							
#molecule_type		DNA							
#residues		1-398 #label FOU2							
#cross-references		EMBL:AJ011856							















THIS PAGE BLANK (USPTO)









CC or send an email to [license@sb-slb.ch](mailto:license@sb-slb.ch)).

Query Match	94%	Score 310	DB 1	Length 433
Best Local Similarity	40.58	Pied. No. 1.10e-24		
Matches 77	Conservative	38	Mismatches 75	Indels 0
				Gaps
Sequence	433 AA	48309 MW	38C4714B	CRC32
				ATP (POTENTIAL)
				AMP-BINDING
				NP-BIND
				114
				120
				433 AA
				48309 MW
				38C4714B
				CRC32
				ATP (POTENTIAL)
				AMP-BINDING
				NP-BIND
				114
				120
				433 AA
				48309 MW
				38C4714B
				CRC32
				ATP (POTENTIAL)
				AMP-BINDING
				NP-BIND
				114
				120
				433 AA
				48309 MW
				38C4714B
				CRC32
				ATP (POTENTIAL)
				AMP-BINDING
				NP-BIND
				114
				120
				433 AA
				48309 MW
				38C4714B
				CRC32
				ATP (POTENTIAL)
				AMP-BINDING
				NP-BIND
				114
				120
				433 AA
				48309 MW
				38C4714B
				CRC32
				ATP (POTENTIAL)
				AMP-BINDING
				NP-BIND
				114
				120
				433 AA
				48309 MW
				38C4714B
				CRC32
				ATP (POTENTIAL)
				AMP-BINDING
				NP-BIND
				114
				120
				433 AA
				48309 MW
				38C4714B
				CRC32
				ATP (POTENTIAL)
				AMP-BINDING
				NP-BIND
				114
				120
				433 AA
				48309 MW
				38C4714B
				CRC32
				ATP (POTENTIAL)
				AMP-BINDING
				NP-BIND
				114
				120
				433 AA
				48309 MW
				38C4714B
				CRC32
				ATP (POTENTIAL)
				AMP-BINDING
				NP-BIND
				114
				120
				433 AA
				48309 MW
				38C4714B
				CRC32
				ATP (POTENTIAL)
				AMP-BINDING
				NP-BIND
				114
				120
				433 AA
				48309 MW
				38C4714B
				CRC32
				ATP (POTENTIAL)
				AMP-BINDING
				NP-BIND











RESULT 8 STANDARD: PRT: 505 AA.

AC MURC\_SYNY3  
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
 DE UDP-N-ACETYLURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-  
 DE ACETYLURAMATE--ALANINE SYNTHETASE).  
 GN MURC OR SL1423.  
 OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).  
 OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 97061201.  
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,  
 RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARDO K., OKUMURA S.,  
 RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,  
 RA TABATA S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RL DNA RES. 3:109-136(1996).  
 CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLURAMATE + L-ALANINE -  
 CC ADP + ORTHOPHOSPHATE + UDP-N-ACETYLURAMATE + L-ALANINE.  
 CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE MURCEP FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D90916; GI653723; -  
 KW PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;  
 KW ATP-BINDING.  
 FT NP\_BIND 164 170 ATP (POTENTIAL).  
 SO SEQUENCE 505 AA; 54513 MW; B3D7FC7C CRC32;

Query Match 8.6%; Score 284; DB 1; Length 505;  
 Best Local Similarity 36.9%; Pred. No. 1,666-20;  
 Matches 120; Conservative 52; Mismatches 150; Indels 3; Gaps 3;

DB A D Y Q A R Q I K R Q A H G N E V E V W  
 DT 861 RGCNCAVYAYCARGCMGNCARATHAARGNCARCNCAVGNAAVAGATGAGTGTG 920  
 DT 13 GGATGACATTTATGCTCAATTTCAATTTGATGATGATGATGATGATGATGATG 72  
 DT D D I Y A Q I F Q I T D K G T A V D V Y  
 DB E R G V C L G T M T V T L P G D H N I S  
 DT 921 GARGNGNGNGTGTGNGNCAATGACNGTACNGTACNGTACNGTACNGTACNGT 980  
 DT 73 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 132  
 DT V D G E F Y D H F L S P Q Y G D H T V L  
 DB N A L A A V A V G R L L G L D F P V I A  
 DT 981 NAAVCNTNG 1040  
 DT 133 AATTCATTTAGCTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 192  
 DT N A L A V I A I S Y L E K L D V T N I K  
 DB Q A I A S F N G A K R R F E C K G Y C N  
 DT 1041 NCARCNATHGNCNGNCAVGNCAVGNCAVGNCAVGNCAVGNCAVGNCAVGN 1100  
 DT 193 AGAAGCATTAAGAAAGTGTGATGATGATGATGATGATGATGATGATGATGAT 251  
 DT E A L E T F G G V K R R F X \* N Y N C K

DB G I T F I D D Y A H H P S E L L A T L A  
 DT 1101 YGNATHACNTTYATHGAYGAYTAAGCAVCAVCAVCAVCAVCAVCAVCAVCA 1160  
 DT 252 ATCAAGT-ATTG-TAGATGATGATGATGATGATGATGATGATGATGATGATG 309  
 DT S S X L X D D Y A H H P R E I S A T I D  
 DB A A K Q K V T H  
 DT 1161 NGCNCNARCARARGTNCNCAY 1185  
 DT 310 CACAGCAGCAAGAAATTCACAT 334  
 DT T A R K K Y P H

RESULT 9 STANDARD: PRT: 483 AA.

AC MURC\_BUCAP  
 DT 15-DEC-1998 (REL. 37, CREATED)  
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
 DE UDP-N-ACETYLURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-  
 DE ACETYLURAMATE--ALANINE SYNTHETASE).  
 GN MURC.  
 OS BUCHNERA APHIDIOLA.  
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; BUCHNERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 98087557.  
 RA BAUMANN L., BAUMANN P.;  
 RT "Characterization of first, the cell division gene of Buchnera  
 RT aphidicola (endosymbiont of aphids) and detection of the product.";  
 RL CURR. MICROBIOL. 36:85-89(1998).  
 CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLURAMATE + L-ALANINE -  
 CC ADP + ORTHOPHOSPHATE + UDP-N-ACETYLURAMATE + L-ALANINE.  
 CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE MURCEP FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF012886; G2738586; -  
 KW PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;  
 KW ATP-BINDING.  
 FT NP\_BIND 124 130 ATP (POTENTIAL).  
 SO SEQUENCE 483 AA; 54852 MW; 23E8697B CRC32;

Query Match 8.1%; Score 266; DB 1; Length 483;  
 Best Local Similarity 38.5%; Pred. No. 1,106-17;  
 Matches 67; Conservative 39; Mismatches 68; Indels 0; Gaps 0;

DB N D Y G H H P T E L S E T I K T I R K S  
 DT 1043 THAAYGATTAAGCAATCAVCAVCAVCAVCAVCAVCAVCAVCAVCAVCA 1102  
 DT 264 TAGATGATTAAGCAATCAVCAVCAVCAVCAVCAVCAVCAVCAVCAVCA 323  
 DT D D Y A H H P R E I S A T I D T A R K K  
 DB W P K K N L I M I F O P H R Y T R T R N  
 DT 1103 SNTGCCNAAARAAAYTTNATGATHTTTTCARCCNCAVMTATACNNGNCA 1162  
 DT 324 AATATCCACATTAAGAGTGTGATGATGATGATGATGATGATGATGATGAT 383  
 DT Y P H K E V V A V F Q P C H T F S R T Q A  
 DB L Y F D F I K I L S O V D S L L I  
 DT 1163 AYYNTATYATGATTTATTAARATHTTNSNCARNGTNGAYMSYTYTNTAT 1216  
 DT 384 CATTTTAAAGATTTGCAAGAAAGTTAGTAAGCAAGATCGTGTATTTAT 437  
 DT F L N E F A E S L S K A D R V V F L











Seq	Sequence	437 AA;	Query Match	100.0%;	Score 3121;	DB 39;	Length 437;
			Best Local Similarity 100.0%;		Pred. No. 5,87e-251;		
			Matches 437; Conservative	0;	Mismatches 0;	Indels 0;	Gaps
Db	1	mtbvhfvgiksgmsiaqinhdiqhewgsgdienyftewalnknkiklprfgannike	60				
QY	1	MTVHFWIGIKSSGSSLSLAQIMHDLGHEVQSGDIEINYFTFVALNNKGIKILPFGANNIKE	60				
Db	61	dmvnylqgnafassheeyirahqkldvvsyndflfgliddqtsvavcgahgkscitglls	120				
QY	61	DMVNYQGNAFSSSHEEYIRAHQLKLDVVSYNDFLFGIIDDQTSVAVGAGHGKSTSTLLS	120				
Db	121	hvmngdkktsfljddqctmgjpsedysfafeaceyrhflsykprdyamtlidfcphdyfk	180				
QY	121	HVMNGDKKTSFLJDDGCTMGJPSEDYFAFEACEYRHRFLSKPRDYAMTINDFCPHDYFK	180				
Db	181	dlnetfdaifgemahnvkkgilawgddehllrkleedvprlyyggfkdsddiyagmqitcdkg	240				
QY	181	DINVEFDFQCMANNVKKGIIAMGDEDEHLRKEIDADVPIYYGGFQDSDDIYAQNQITDKG	240				
Db	241	tafdvyvqgefyvdhflspqgygdhvlmalavialasyekidvtrikealtctfgvkrfrfn	300				
QY	241	TAFDVYVDGEFYDHFLSPOYGDHVLNALAVIALASYEKIDVTRIKALETFEGGVKFRFN	300				
Db	301	ettlangyivddyahhpnejatidtarckkypkhevavfgphtfstrtqaflnofaeslc	360				
QY	301	ETTIANQYIVDDYAHHPNEJATIDTARKKTPKHEVAVAVQPHFSTRTOAFLNFFAESLC	360				
Db	361	kadvrficelafgsirensгалttdqldldklygnasfinedlinvleqfdanavlfmgagdi	420				
QY	361	KADVRFICEIGFSIRENSGALTIDQLDLKIGGASFINEDLINVLEQDNNAVLFMAGADI	420				
Db	421	qklnqneyajdklgmknaf	437				
QY	421	QKLNQNEYAJDKLGKMNAF	437				

ID	RESULT
AC	W89199 standard; Protein; 215 AA.
AD	W89199;
DE	17-MAR-1999 (first entry)
DT	Partial sequence of the MucR polypeptide.
KW	MucR gene: UDP-N-acetylmuramate:L-alanine ligase; MucR polypeptide;
KW	bacterial: infection; H. pylori; cancer; ulcer; gastritis; vaccine;
KW	immunogen; drug; genetic immunisation.
OS	Staphylococcus aureus.
PN	EP-889123-A2.
PD	07-JAN-1999.
PF	26-JUN-1998; 305064.
PR	03-JUL-1997; US-052720.
PA	(SMIK ) SMITHKLINE BEECHAM CORP.
PI	Burnham MKR, Wallis NG;
DR	WPI: 99-062655/06.
DR	N-PSDB: V80065.
PT	New isolated MucR polypeptide from Staphylococcus aureus and related
PT	nucleic acid - useful in diagnosis, treatment and prevention of
PT	bacterial infections
PS	Claim 1; Page 5; 39pp; English.
CC	The invention relates to a UDP-N-acetylmuramate:L-alanine ligase
CC	(MucR polypeptide) encoded by the S. aureus MucR gene. Host cells
CC	containing an expression system comprising the MucR gene can be used for
CC	the recombinant production of the polypeptide. Agonists or the MucR
CC	polypeptide are used to treat conditions requiring increased activity or
CC	expression of the polypeptide. Antagonists, inhibitory nucleic acid or
CC	competitive polypeptide are useful for inhibiting the polypeptide e.g.
CC	bacterial (especially S. aureus) infections. They are also useful against
CC	Helicobacter pylori infections and related cancers, ulcers and gastritis
CC	The antibacterial agents are useful to treat in-dwelling devices for
CC	infection prevention or generally as wound treatments to prevent adhesion
CC	of bacteria to matrix proteins. The MucR polypeptide is also useful for

CC diagnosing or prognosing a (susceptibility to) disease, for raising  
CC antibodies; to identify modulators or specific receptors, in rational  
CC drug design and as an immunogen for vaccines. The MucC gene sequences are  
CC useful in antisense/ribozyme therapeutics; to detect mutant MucC gene;  
CC for chromosomal mapping; to determine bacterial serotype; and for genetic  
CC immunisation. The present sequence represents a partial sequence of the  
CC MucC polypeptide.  
CC .Sequence 215 AA;  
CC 90

Query Match	48.3%	Score 1509	DB 39	Length 215
Best Local Similarity	100.0%	Pred. No. 9.35e-114		
Matches 215	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Db 1	fkdsddiaqagiqitdkgtaktdvdyvdgefydhflspgyghvtnalavaiasyekldv	60		
Oy 223	FKSDDDIAYQNIQTDRGTAFDVYVDGEFYDHFSLPGYGHVTLNALAVIAISYLERLDV	282		
Db 61	tnlkealeifggyvrrfnettianqvivddyahpraisitidatrkkyphkewavfqp	120		
Oy 283	TNlKEAlEtFGvYRRfNETtIANQvIVdDYAHhPRISITIdATrKkYPhKEwAVFQp	342		
Db 121	htsfirtqafineafesjckadvrfelceifgsirensagatitqglidkiggasfinegl	180		
Oy 343	HTSFRTQAFInEFAfESjCKADrvFELCEIfGSIReNSAGAlTQGLIdKIGASfInEDlN	402		
Db 181	vleqfdhnavvlfmgagdtqrlqmayidkismknaif	215		
Oy 403	VLEQFDhNAVvLFMGAGDlTQRLQmAYIdKISmKNAIf	437		

RESULT 3  
 ID W55120 standard; Protein; 422 AA.  
 AC W55120;  
 DT 02-Oct-1998 (first entry)  
 DE Streptococcus pneumoniae SP0070 protein.  
 KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
 RN detection; pneumonia; otitis media; meningitis.  
 SN Streptococcus pneumoniae.  
 PS M09818930-A2.  
 PN 07-MAY-1998.  
 PD 30-Oct-1997; U19422.  
 PR 31-Oct-1996; US-029960.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Choi GH, Hromocky A, Johnson LS, Kunsch CA;  
 DR WPI: 98-272224/24.  
 DR N-PSDS: V27381.  
 PT Nucleic acid encoding antigenic peptide(s) from Streptococcus  
 PT pneumoniae - or their epitope-containing fragments, useful in  
 PT protective or therapeutic vaccines, and for diagnosis  
 PS Claim 11; Page 73; 118PP; English.  
 CC The present sequence represents a protein from Streptococcus pneumoniae  
 CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
 CC can be useful in vaccines for inducing protective antibodies against  
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
 CC are used to detect Streptococcus infection (by usual hybridisation or  
 CC amplification methods), also for isolating Streptococcus genes or their  
 CC allelic variants. The protein can be used similarly to detect specific  
 CC antibodies in standard immunoassays, especially for diagnosing or  
 CC monitoring infections. Antibodies which bind the protein are used to  
 CC detect corresponding antigens, to purify the protein and for passive  
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
 CC (especially 10-300) mu g/ml per dose.  
 SE Sequence 422 AA:

	Query Match	47.2%	Score 1473;	DB 33;	Length 422;
	Best Local Similarity	48.6%;	Pred. No. 1,03e-110;		
	Matches 201; Conservative	94;	Mismatches 114;	Indels 5;	Gaps 5;
Db	1 hgmghkvygsdvckkyftcrlglegagtltpfcdknldgdmelagnafipdnvvelaya	60			
	:::::     :::     :::     :::     :::     :::     :::     :::				
xy	22 HDLHEVQGSIDIEENVYEVLEALNRKKGIRKLPEFGNNNTKEDMYVLOGAFASSNE-ETVRA	80			



\*\*\*\*\*  
MUSE RELEASE  
\*\*\*\*\* (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msrch\_p protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Nov 24 00:38:50 1999; Maspar time 17.41 Seconds

Tabular output not generated. 533.812 Million cell updates/sec

Title: >US-09-103-287-2  
Description: (1-437) from US09103287.pep  
Perfect Score: 3121  
Sequence: 1 MTHYHFVIGKSGSMSSLAQI.....GDIQKQNAIYDILGKKNAF 437

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 34.980; Variance 176.086; scale 0.199

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	3121	100.0	437	39	W87771	5.87e-251
2	1509	48.3	215	39	W89199	9.39e-114
3	1473	47.2	422	33	W55120	1.03e-110
4	405	13.0	455	22	W20606	6.99e-22
5	273	8.7	46	34	W77686	1.70e-11
6	239	7.7	286	22	W24585	6.71e-09
7	239	7.7	286	22	W20102	6.71e-09
8	143	4.6	450	26	W29454	6.65e-02
9	137	4.6	450	37	W68551	6.65e-02
10	137	4.4	335	33	W55117	1.72e-01
11	128	4.1	481	26	W34453	6.97e-01
12	118	3.8	1786	23	W24790	3.20e+00
13	111	3.6	370	35	W71543	9.12e+00
14	110	3.5	536	18	W01107	1.06e+01
15	110	3.5	1670	18	R37048	1.06e+01
16	110	3.5	1670	16	R85599	1.06e+01

17	110	3.5	1670	23	W20049	KOD1 thermostable DNA	1.06e+01
18	110	3.5	1670	28	W41314	Full length heat-resi	1.06e+01
19	110	3.5	1670	18	R94616	DNA polymerase.	1.06e+01
20	106	3.4	432	22	W20733	H. pylori cell envelo	1.90e+01
21	105	3.4	806	5	R27481	RP-III residual prote	2.20e+01
22	104	3.3	187	22	W20436	H. pylori protein.	2.54e+01
23	103	3.3	231	3	R12878	N-acetylglucosamine d	2.94e+01
24	102	3.3	3060	23	W22475	Plasmodium var-7.	3.39e+01
25	99	3.2	400	22	W20826	H. pylori cytoplasmic	5.21e+01
26	97	3.1	62	22	W20505	H. pylori cytoplasmic	6.91e+01
27	98	3.1	112	31	W27907	Amino acid sequence o	6.00e+01
28	97	3.1	165	22	W20621	H. pylori cytoplasmic	6.91e+01
29	96	3.1	283	29	W55453	H. pylori ORF 05ae302	7.96e+01
30	96	3.1	293	29	W55244	H. pylori ORF 05ae202	7.96e+01
31	98	3.1	422	22	W20249	H. pylori transmembra	6.00e+01
32	98	3.1	426	22	W20934	H. pylori surface or	6.00e+01
33	98	3.1	510	33	W61269	Helicobacter pylori f	6.00e+01
34	97	3.1	1019	8	R38096	Staphylococcus sp DNA pol	6.91e+01
35	97	3.1	1019	16	R81554	Pyrococcus sp. DNA po	6.91e+01
36	97	3.1	1312	37	W70563	DNA polymerase of a	6.91e+01
37	97	3.1	1312	10	R53787	Deep Vent DNA polymer	6.91e+01
38	95	3.0	110	22	W20533	H. pylori cytoplasmic	9.16e+01
39	94	3.0	345	16	R77026	Canine herpesvirus gd	1.05e+02
40	94	3.0	345	36	W72658	Canine herpes virus p	1.05e+02
41	94	3.0	345	26	W22998	Canine herpesvirus g1	1.05e+02
42	94	3.0	411	21	W01558	Human CD36 antigen.	1.05e+02
43	94	3.0	472	4	R20810	Streptococcus pneumo	1.05e+02
44	94	3.0	496	31	W55962	Secreted protein c1on	1.05e+02
45	94	3.0	677	38	W85608		

#### ALIGNMENTS

RESULT 1  
ID W87771 standard; Protein; 437 AA.  
AC W87771;  
DE UDP-N-acetylmutamate:L-alanine ligase (MucC polypeptide).  
KW MucC gene; UDP-N-acetylmutamate:L-alanine ligase; MucC polypeptide;  
KW bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
KW immunogen; drug; genetic immunisation.  
OS Staphylococcus aureus.  
PN EP-889123-A2.  
PD 07-JAN-1999.  
PE 26-JUN-1998; 305064.  
PR 03-JUL-1997; US-052720.  
PA (SMIK ) SMITHKLINE BECHAM CORP.  
PI Burnham MKR, Wallis NG;  
DR WPI; 99-062655/06.  
DR N-PSDB; V99650.  
PT New isolated MucC polypeptide from Staphylococcus aureus and related  
PT nucleic acid - useful in diagnosis, treatment and prevention of  
PT bacterial infections  
PS Claim 1; Page 4; 39pp; English.  
CC The present sequence represents a UDP-N-acetylmutamate:L-alanine ligase  
CC (MucC polypeptide) encoded by the S. aureus MucC gene. Host cells  
CC containing an expression system comprising the MucC gene can be used for  
CC the recombinant production of the polypeptide. Agonists or the MucC  
CC polypeptide are used to treat conditions requiring increased activity or  
CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or  
CC competitive polypeptide are useful for inhibiting the polypeptide e.g.  
CC bacterial (especially S. aureus) infections. They are also useful against  
CC Helicobacter pylori infections and related cancers, ulcers and gastritis.  
CC The antibacterial agents are useful to treat in-dwelling devices for  
CC infection prevention or generally as wound treatments to prevent adhesion  
CC of bacteria to matrix proteins. The MucC polypeptide is also useful for  
CC diagnosing or prognosing a (susceptibility to) disease, for raising  
CC antibodies to identify modulators or specific receptors, in rational  
CC drug design and as an immunogen for vaccines. The MucC gene sequences are  
CC useful in antisense/ribozyme therapeutics, to detect mutant MucC gene;  
CC for chromosomal mapping; to determine bacterial serotype; and for genetic  
CC immunisation.







## FEATURES

Seq primer: -41m13 fwd. ET from Amersham  
High quality sequence stop: 245.  
Location/Qualifiers  
1..292

## BASE COUNT

103 a 62 c 40 g 87 t

## ORIGIN

## Query Match

Best Local Similarity 76.6%; Pred. No. 6.31e-04;  
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 127 GATCTAATTGAGAAATGAAAGATATCTGCTACATAAGACATCT 173

1174 GATTATATGATTAATAATGGAGCTGATCTTATTAATGAAGATCT 1220

## RESULT 14

LOCUS AA041443 296 bp mRNA EST 01-FEB-1997  
DEFINITION zifog905.r1 Soares fetal heart\_NbHH19W Homo sapiens cDNA clone  
IMAGE:376472 5', mRNA sequence.

## ACCESSION

AA041443

NTD 91517668

VERSION AA041443.1 GI:1517668

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 296)

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, R., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisan, E.,

Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.

WashU-Merck EST Project

Unpublished (1995)

On Nov 29, 1993 this sequence version replaced gi:430399.

TITLE

JOURNAL

COMMENT

## FEATURES

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 1196 Std Error: 0.00  
Seq primer: -28m13 rev2 from Amersham  
High quality sequence stop: 259.  
Location/Qualifiers  
1..296  
/organism="Homo sapiens"  
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st

## BASE COUNT

66 a 74 c 68 g 87 t 1 others

## ORIGIN

## Query Match

Best Local Similarity 71.2%; Pred. No. 6.31e-04;  
Matches 42; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Db 115 CTGTGCTTCAGCTGTCGAGATGAATGCTGTTACCTGCATTCATGTTTAT 173

433 CTGTGCTTCAGCTGTCGAGATGAATGCTGTTTATCACCATTCATGATGATAT 375

## RESULT 15

LOCUS AA229413 307 bp mRNA EST 21-AUG-1997  
DEFINITION nc47f10.r1 NCI\_CGAP\_P3 Homo sapiens cDNA clone IMAGE:1011307, mRNA  
sequence.

## ACCESSION

AA229413

NTD 91851266

VERSION AA229413.1 GI:1851266

## KEYWORDS

EST.

## SOURCE

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 307)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

On Sep 12, 1996 this sequence version replaced gi:1395052.

JOURNAL

COMMENT

## FEATURES

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.lnl.gov/db/ftp/image/image.html  
Insert Length: 388 Std Error: 0.00  
Seq primer: -28m13 rev1 ET from Amersham.  
Location/Qualifiers  
1..307  
/organism="Homo sapiens"  
/note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected cells  
histologically-determined to be fully malignant prostate  
cancer cells. Double-stranded cDNA was ligated to EcoRI  
adaptors, 5 cycles of PCR applied to the cDNA with an  
adaptor-specific primer, and the resulting PCR product



ATTTAATTCATAAATGGAGGTGCATCGTTCATTAAIGAGATCT 1220

This clone is avail



Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/ldb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13 Reverse  
Class: BAC ends.

## FEATURES

## Source

Location/Qualifiers  
1..694

/organism="Homo sapiens"  
/note="Vector: pBelBAC11; Site\_1: HindIII; Site\_2:  
HindIII"  
/db\_xref="taxon:9606"  
/clone="238N10"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"

BASE COUNT 250 a 107 c 96 g 241 t  
ORIGIN

## Query Match

Best Local Similarity 76.5%; Score 27; DB 38; Length 694;  
Pred. No. 4,24e-06;  
Matches 39; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Db 487 ATTCTATTATGTCATTGACACATGATGTAATGTTGTAATTAAGA 537  
111111 111111 111111 111111 111111 111111 111111 111111 111111 111111  
Cp 596 ATTCTTGATGCATCAAAAACATCATTAATATTCTTGAAATTAACAGA 546

RESULT 9  
LOCUS AA090223 238 bp mRNA EST 24-OCT-1996  
DEFINITION cchp0673.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens  
ACCESSION AA090223  
NID 91636707  
VERSION AA090223.1 GI:1636707  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 238)  
AUTHORS Liew,C.C.  
TITLE CDNAs from fetal heart (1996)  
JOURNAL Unpublished (1996)  
COMMENT On Jan 24, 1995 this sequence version replaced gi:634334.

Contact: Liew CC  
Department of Laboratory Medicine and Pathobiology  
University of Toronto  
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5  
Tel: 4169788758  
Fax: 4169785650  
Email: liewcc@utcc.utoronto.ca  
PCR Primers  
FORWARD: 5' GCCAGCTCGAATTAACCTCACTAAAGG 3'  
BACKWARD: 5' CCACTGATGTATAGACACTCACTATGCGC 3'  
Seq primer: 5' GAATTAACCTCACTAAAGG 3'.  
Location/Qualifiers  
1..238

## FEATURES

## Source

/organism="Homo sapiens"  
/note="Vector: Lambda ZAP Express; Site\_1: EcoRI; Site\_2:  
XhoI; mRNA was purified from human fetal hearts (8-10  
weeks). cDNA was synthesized using a XhoI-Oligo dT  
adaptor-primer. EcoRI adaptors were ligated, followed by  
digestion with XhoI, for directional cloning into  
predigested lambda ZAP Express."  
/db\_xref="taxon:9606"

/clone\_lib="Human fetal heart, Lambda ZAP Express"  
/lab\_host="E. coli XL1-Blue"  
BASE COUNT 61 a 60 c 48 g 69 t  
ORIGIN

Query Match 1.9%; Score 25; DB 36; Length 238;  
Best Local Similarity 71.2%; Pred. No. 6,31e-04;  
Matches 42; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Db 99 CTGTGCTTCAGCTGTGAGATGAATGTTCTGTACCTCGATTAACATGTTAT 157  
111111 111111 111111 111111 111111 111111 111111 111111 111111 111111  
Cp 433 CTGTGCATCACCATTAAATGAAGTCTTTTATCACCATTCAATACATGTGATAT 375

RESULT 10  
LOCUS AA229467 277 bp mRNA EST 21-AUG-1997  
DEFINITION nc45f02.r1 NCI\_CGAP\_Pt3 Homo sapiens cDNA clone IMAGE:1011099, mRNA  
sequence.  
ACCESSION AA229467  
NID 91851302  
VERSION AA229467.1 GI:1851302  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 277)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1394311.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILMIL at:  
www.bio.lnl.gov/dbtrp/image/image.html

Insert Length: 448 Std Error: 0.00  
Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 256.  
Location/Qualifiers  
1..277

## FEATURES

## Source

/organism="Homo sapiens"  
/note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected cells  
histologically-determined to be fully malignant prostate  
cancer cells. Double-stranded cDNA was ligated to EcoRI  
adaptors, 5 cycles of PCR applied to the cDNA with an  
adaptor-specific primer, and the resulting PCR product  
subcloned into pAMP10 by the UDG-cloning method (Life  
Technologies). Average insert size is 600 bp. NOTE: Not  
directionally cloned. This library was constructed by  
David Krizman."

/db\_xref="taxon:9606"  
/clone="IMAGE:1011099"  
/clone\_lib="NCI\_CGAP\_Pt3"  
/sex="Male"  
/dev\_strage="45 years old"  
/lab\_host="DH10B"

BASE COUNT 94 a 44 c 44 g 95 t  
ORIGIN  
Query Match 1.9%; Score 25; DB 10; Length 277;



Db 1532 RRYKRYTSTYKSMRWYTTTYTMYCWCCTSMK-SASACAMRMWYMGSSRSRSGY 1590  
 Cp 792 AGCACGAGAGTGTCTAATAAAGTACCATCATACATCAAGAGAGTCTTATC 733  
 Db 1591 WGSMSGCGYMTKRYRYTGTWTTWYMWMSMTWMTTWTWTTWTTKTTMMWMT 1650  
 Cp 732 CGTAATTGAATATTGTGACATAAATGATCCGATCTTTAAATCAATAGTAATAAT 673  
 Db 1651 T 1651  
 Cp 672 T 672

RESULT 6  
 LOCUS AF034173 2275 bp mRNA EST 30-MAR-1998  
 DEFINITION AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens CDNA  
 clone ntcon2 contig, mRNA sequence.  
 ACCESSION AF034173  
 NID 92707735  
 VERSION AF034173.1 GI:2707735  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 2275)  
 AUTHORS Tripodis,N. and Ragoussis,J.  
 TITLE Generation of a transcription map in the region immediately  
 centromeric to human MHC across the 6p21.2-6p21.3 chromosomal  
 boundary  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2045115.

FEATURES  
 source Location/Qualifiers  
 1..2275  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="6p21.3"  
 /clone="ntcon2 contig"  
 /clone\_lib="Human mRNA (Tripodis and Ragoussis)"  
 BASE COUNT 438 a 619 c 470 g 599 t 149 others  
 ORIGIN

Query Match 2.7%; Score 36; DB 20; Length 2275;  
 Best Local Similarity 13.5%; Pred. No. 4.23e-17;  
 Matches 26; Conservative 98; Mismatches 64; Indels 4; Gaps 4;

Db 1479 WRYKRRKRRGRRRRTGMYKRYRAMMACAMACWYWRKMKCKYK-YKRY 1537  
 Oy 709 TATGCTCAAAATTCATTAAGGATGAAGGACTGCTTTGATGATGATGG 768  
 Db 1538 T-STYKSMRKWTTTYTYWCCTSMKSASACAMRMWYMGSSRSRSGY 1596  
 Oy 769 GAGTTTATGATCATCTCTGCTCCACATATGATGACCATTA-CAGTTTAAATGCATT 827  
 Db 1597 C-YGMYKRYRYSWTGTWTTWYMWMSMTWMTTWTWTTWTTKTTMMWMT 1655  
 Oy 828 AGCTGTAATTCGATTAATTAAGAGAGCTAGATGTTACAAATTAAGAGACATT 867  
 Db 1656 RRRGAGTTTGG 1667  
 Oy 888 AGAAGCTTTGG 899

RESULT 7  
 LOCUS B50349 439 bp DNA GSS 19-JUN-1998  
 DEFINITION CIT-HSP-351C15.TV CIT-HSP Homo sapiens genomic clone 351C15,

genomic survey sequence.  
 B50349  
 NID 92602586  
 VERSION B50349.1 GI:2602586  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 439)  
 AUTHORS Kim,U.-J., Adams,M.D. and Simon,M.I.  
 TITLE Determination of clone end sequences of human Bacterial Artificial  
 Chromosomes  
 JOURNAL Unpublished (1997)  
 COMMENT Other\_GSSs: CIT-HSP-351C15.TP  
 Contact: Ung-Jin Kim  
 Caltech Genome Research Lab  
 California Institute of Technology  
 Division of Biology, MS 147-75, Pasadena, CA 91125, USA  
 Tel: 626 796 7066  
 Fax: 626 395 4901  
 Email: ung@ash-tree.caltech.edu  
 Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:  
 http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html  
 Seq primer: 77  
 Class: BAC ends.

FEATURES  
 source Location/Qualifiers  
 1..439  
 /organism="Homo sapiens"  
 /note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
 HindIII"  
 /db\_xref="taxon:9606"  
 /clone="351C15"  
 /clone\_lib="CIT-HSP"  
 /sex="Male"  
 /cell\_type="Sperm"  
 BASE COUNT 108 a 115 c 75 g 141 t  
 ORIGIN

Query Match 2.1%; Score 28; DB 38; Length 439;  
 Best Local Similarity 72.6%; Pred. No. 3.14e-07;  
 Matches 45; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Db 209 ATGCATAGAGGATCTCTAGATATCATCCACCCATTAATATACACGAATTTGAA 268  
 Cp 1300 ATGCATTTGTAATTTTGAATATACCTGCACCCATTAATTAACACAGCATATCA 1241  
 Db 269 AT 270  
 Cp 1240 AT 1239

RESULT 8  
 LOCUS A0056084 694 bp DNA GSS 29-JUL-1998  
 DEFINITION CIT-HSP-2338N10.TR CIT-HSP Homo sapiens genomic clone 2338N10,  
 genomic survey sequence.  
 ACCESSION A0056084  
 NID 93352690  
 VERSION A0056084.1 GI:3352690  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 694)  
 AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,  
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,  
 Simon,M. and Venter,J.C.  
 TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map  
 Building  
 JOURNAL Unpublished (1997)  
 COMMENT Other\_GSSs: CIT-HSP-2338N10.TF



```

/cultivar="Milyang23"
/node=vector: PBInescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZapII
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1784"
/clone_lib="Rice Immature Seed Lambda ZapII cDNA library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"

```

	Query Match	3.5%	Score 47	DB 17	Length 247
	Best Local Similarity	12.4%	Pred. No. 2,47e-32		
	Matches	26	Conservative	99	Mismatches 82, Indels 2, Gaps 2;
Db	19	WMNKHMTBMCVRRVCGTTTNGKHNGRPTTWDCSDNAHCRTYVEMYARSKYGG	78		
Oy	730	ACGGTAAAGCTAGCTGTTTGATGTGTAGTACGATGAGCTTATGATCACTTCG	789		
Db	79	TBYYSAMNDVNTGCTGKTVVNHSHSMNRCSSVYVYBMYCDBHYEDRAHHVDY	138		
Oy	790	TCTCCACATATGGTGACCATTA-CAGTTTAAATGCTTAGCTGATATGCCATTAGTA	848		
Db	139	RCINDRGICNYTASDNGSTSAIKRYTGDKTDSOCGGCWMKRYVYSSBIBRCGVNMYRT	198		
Oy	849	TTTAGAGACGCTAGATGTTCACAAATATTAAAGAAGCATTTGAACCGTTGGTGCTGTAA	908		
Db	199	TSMNTDKSTKM-BSMDSRSRSRYVYGRMM	226		
Oy	909	ACGTGGTTCATGAACTACAAATTTGCA	937		

RESULT	LOCUS	DEFINITION	ACCESSION	NID	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT			
4	AA754458	247 bp mRNA, EST	97SN1784	Rice Immature Seed Lambda ZAPII cDNA Library	Oryza sativa	AA754458	92801164	GI:2801164	EST.	Oryza sativa.	Oryza sativa	Enkaryote; Viridiplantae; Streptophyta; Embryophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza	1 (bases 1 to 247)	Nahm, B.H., Kim, J.K., Cheong, J.J., Kim, S.I., Hahn, T.R., Moon, E.P., Kim, W.T., Kim, W.Y., Yang, M.S., Park, R.D., Sohn, U.I., Kang, K.Y., Lee, M.C. and Eun, M.Y.	large-scale Sequencing Analysis of ESTs from Rice Immature Seed Unpublished (1998)	On Jan 14, 1998 this sequence version replaced gi:1797455.

**CONTACT:** Eun M.Y.  
 Department of Cytogenetics  
 National Inst. of Agri. Sci. and Tech, RDA  
 Suwon, Kyunggi-do, Korea  
 Tel : 82 331 290 0301  
 Fax: 82 331 290 0307  
 Email: myeun@sun20.asstl.re.kr  
 Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji  
 University, Yongin, Korea, 449-728 bminah@bloserver.myongji.ac.kr  
**Seq primer:** M13 Reverse Primer.  
**Location/Qualifiers**  
 1. 247  
 /organism="Oryza sativa"  
 /cultivar="Milyang23"  
 /note="Vector: pBluescript SK(+), Site\_1: EcoRI, Site\_2:  
 XhoI, Directional CDNA library inserted into lambda ZapIII  
 vector at 5' end with EcoRI and 3' end with Xho I site."

```

/db_xref="taxon:4530"
/map="6"
/clone="97SN1784"
/clone_lib="Rice Immature Seed Lambda ZapII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SDR"
BASE COUNT      7 a      16 c      21 g      34 t      169 others
ORIGIN

Query Match      3 4%; Score 46; DB 17; Length 247;
Best Local Similarity 14.7%; Pred. No. 6,86e-31;
Matches 33; Conservative 104; Mismatches 84; Indels 4; Gaps 4;

```

Db 22 KTHMHTBMCVRRVGTGTTNGNKHGRITTNWDCSDNHNCRITVBMYIARSKYGGTBY 81  
Cp 1146 TCTAATGTGGCCAAAATTTACATAGATACACGATCT-GCTTTACATAACTT-TCT 1089  
Db 82 YSMNDVTNGCTGCVKTTVNVHSGWNRRCSNSVYVWETAYCYDBYBDBRANHVDTRCT 141  
Cp 1088 GCAATTCATTTAAAAAG -CTTGCTGTTCTAGAGAAAGTGCTGCTGAATTCCTGCAC 1030  
Db 142 NDRGVCNTA -SDNGTSATKRVGTGDKTDDCGGCGMKRVYTGSSBYBRGCVNVMRTTS 200  
Cp 1039 AACTCTTTATCTGGATATTTCTTCGCGCTGTGCAATGTAGCACTAATTTCTCTGG 970  
Db 201 MWTDSTKMBSDMRSRSHVHYGRMBNKKRQMSRWEDTCKMRT 245  
Cp 969 ATGGGTGGCATATCATCTACATAACTTGATTTGGCAATTTGAT 925

RESULT	5				
LOCUS	AF034173	2275 bp	mRNA	EST	30-MAR-1998
DEFINITION	AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens CDNA clone ntcon2 contig, mRNA sequence.				
ACCESSION	AF034173				
NID	G2707735				
VERSION	AF034173.1	GI:2707735			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Primates; Catarrhini; Homidae; Homo.				
REFERENCE	1 (bases 1 to 2275)				
AUTHORS	Tripodis,N. and Ragoussis,J.				
TITLE	Generation of a transcription map in the region immediately centromeric to human MHC across the 6p21.2-6p21.3 chromosomal boundary				
JOURNAL	Unpublished (1997)				
COMMENT	On Jan 19, 1998 this sequence version replaced gi:2045115.				

<b>FEATURES</b>						
		<b>SOURCE</b>				
		<b>Location/Qualifiers</b>				
		1.	.2275			
		/organism="Homo sapiens"				
		/db_xref="taxon:9606"				
		/map="6p21.3"				
		/clone="ntcon2 contig"				
		/clone_lib="Human mRNA (Tripodis and Ragoussis)"				
BASE COUNT		438 a	619 c	470 g	599 t	149 others
ORIGIN						
Query Match		2.9%	Score 39;	DB 20;	Length 2275;	
Best Local Similarity		12.7%;	Pred. No. 4.24e-21;			
Matches	23; Conservative			Mismatches 64;	Indels 2;	Gaps 2;
Db	1472 AAAAAAAAAA	RKKRKRKRRKRRKRRMTGTCRTKRYTAAAMACMACAACWCYYTKMKMGKKCKMY	1531			
p	851 AAATCACTAATTCGCAGATTACAGCTAATCCAGTTTAACAACGTATGGTC -ACCATAATGTGG	793				



```

/cultivar="Milyang33"
/notes=Vector: pBluescript SK(+), Site_1: EcoRI, Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPIT
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1787"
/clone_1b="Rice Immature Seed Lambda ZAPIT cDNA library"
/dev_tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"

```

	Query Match	5.0%	Score 67	DB 17	Length 252
	Best Local	Similarity 15.6%	Pred. No. 8,55e-63		
	Matches	39	Conservative	126	Mismatches 79; Indels 6; Gaps 6;
Db	3	DCIATNATYAGCCCCCAAAATTTSYBCHGNEBWWCYAASHGNTYSVHNCIBRGTCHCCKKANNVNST	62		
Cp	1146	TCTATATGAGCAAAATTTCCACATAAGATAACAGCATCTGCTTATACATAAACTTCTGC	1087		
Db	63	MTATGTVMBANVSCDMH-YMEVBNTKVDVGNHTRCSRMRYBTAHAYHDYTNBRYNNNDY	121		
Cp	1086	AAATTCATTTAAANAATGCTTGTTGTAG-AGANAAGTGTGTGTTGAATTAATCTGCACAA	1028		
Db	122	HHMBBMBTBEGTCCTTCMCMBYHNKCTASGHHSTINDVKS-SINTYGVATBSIDKSMH	180		
Cp	1027	CTTCATTATGAGATTTCTTCCTGCTGCTGTCAATTGTAGACACTAATTTCTTGGAT	968		
Db	181	GYMCSSBAYKYTKV-STT-RATRSYTCVARKYCYMMATKKYVHVBBGCH-BTDSCK	237		
Cp	967	GGTGCGCATATCATCTACAAATAACTGATTTGCATTTGAGTTTCATTTGAACAGCGTT	908		
Db	238	TMMNTNKHYM	247		
Cp	907	TAAACACACC	898		

[illegible]

Contact: Eun M.Y.  
 Department of Cytoogenetics  
 National Inst. of Agri. Sci. and Tech, RDA  
 Suwon, Kyunggi-do, Korea  
 Tel.: 82 331 290 0301  
 Fax: 82 331 290 0307  
 Email: myeun@sunt0.asri.re.kr  
 Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji  
 University, Yongin, Korea, 449-728 bhnaem@bserver.myongji.ac.kr  
 Seq primer: M13 Reverse Primer.  
 Location/Qualifiers  
 1. .252  
 /organism="Oryza sativa"

```

/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZapII
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SM1787"
/clone_lib="Rice Immature Seed Lambda ZapII cDNA library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"

```

BASE COUNT	ORIGIN
5 a	21 c
12 g	35 t
179 others	

	Query Match	4.0%	Score 54;	DB 17;	Length 257;
	Best Local Similarity	12.4%	Pred. No.	1,11e-42;	
	Matches	30;	Conservative	117;	Mismatches 92; Indels 3; Gaps 3;
Dd	8	TVRGCCCBAMTTT SYCHGNBYWVCVASHGNTMSVHNCBTBGRTHODCKNVKMTMGT	67		
Oy	1041	TCACACACACTTCTCTCAGAACACAAGCATTTTAATGAATTGGCAAAAGTTA-T	1099		
Dd	68	VNMENVSGDMHYWVBVTKYDVGNHTRCSSMRPVTMAHYHDYTCNBRYNNNDYMIBB	127		
Oy	1100	GTAAGCGAAGTGCGTTCCTTATGTGAATTTTGGCCATTTAGAG-AAATCTCTGC	1158		
Dd	128	MYBTGCCTCTMCMBHYNTKCTASGMHSTINDVKSSTNITMGVLTBSIDKSMHGWCBB	187		
Oy	1159	GCATTAACGATACAAAGATTTAATGTATAAATGGAGGTCATCCTTATTATTAAGAAT	1218		
Dd	188	VKHVTKYST-TFRATRSYCVRKCYVMNMTRKKVKYKVHVBCSHDSCDKTMWTNKHV	246		
Oy	1219	CTFTATGATGATTAGAACAAATTTGAATATGCGTGTGTTTATTTATGGGTCACAGTAT	1278		
Dd	247	MT 248			
Oy	1279	AT 1280			

RESULT	3								
LOCUS	AA754458	247 bp	mRNA	EST	20-JAN-1998				
DEFINITION	g7SN1784 Rice Immature Seed Iamdba ZAP II cDNA Library Oryza sativa								
ACCESSION	cDNA clone g7SN1784, mRNA sequence.								
NID	AA754458								
VERSION	g2801164								
KEYWORDS	AA754458.1	GI:2801164							
SOURCE	EST.								
ORGANISM	Oryza sativa.								
AUTHORS	Oryza sativa Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eumphylophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza. 1 (bases 1 to 247)								
REFERENCE	Nahn,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,M.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y. Large-scale Sequencing Analysis of ESTs from Rice Immature Seed Unpublished (1998) On Jan 14, 1998 this sequence version replaced gi:1797455.								
TITLE									
JOURNAL									
COMMENT									

**CONTACT:** Eun M. Y.  
Department of Cytogenetics  
National Inst. of Agri. Sci. and Tech., RDA  
Suwon, Kyunggido, Korea  
Tel.: 82 331 290 0301  
Fax: 82 331 290 0307  
Email: myeun@sun20.asti.re.kr  
Submitted by Baek Hie Nam, Dept of Biological Science, Myongji  
University, Yongin, Korea. 449-728 bhnahe@bioserver.myongji.ac.kr  
Seq primer: M13 Reverse Primer.  
location/Qualifiers  
1. 247  
/organism="Oryza sativa"







**THIS PAGE BLANK (USPTO)**



RA VAN VUOT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,  
RA UTTERBACK T., WATHEY L., McDONALD L., ARTACH P., BOWMAN C.,  
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,  
RA SMITH H.O., VENTER J.C.;  
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia  
burgdorferi";  
RL NATURE 380:580-586(1997).  
DR EMBL: AE000791; G2689913; -.  
DR TIGR: BRC01; -.  
KW PLASMID.  
SQ SEQUENCE 369 AA; 44819 MW; AAB2E7C8 CRC32;

Query Match 2.5%; Score 170; DB 2; Length 369;  
Best Local Similarity 44.3%; Pred. No. 8.80e-03;  
Matches 43; Conservative 23; Mismatches 31; Indels 0; Gaps 0;

Db F L T I L F N L D I D K D N I I K I L K  
Dt 534 HTTYTACNATHTNTTYAAYTNGAATGAAAGAAAYATHTAARATHTYNA 593  
Cp 616 CTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 557  
Ct F N I M C H F L E C I K N I I N I F E  
Db I I K I I E I K L L K N  
Dt 594 RATHAATATHTHATGAAATHTATATATATATATATATATATATATATAT 630  
Cp 556 AATATCAGATGATCGAATCAATATATATATATATATATATATATATATAT 520  
Ct I I R M I E I N I C H N

RESULT 14  
ID Q25768 PRELIMINARY; PRT; 391 AA.  
AC Q25768;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE ASPARAGINE-RICH ANTIGEN (CLONE 14C1) (FRAGMENT).  
OS PLASMODIUM FALCIPARUM.  
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SCHREIBER M.G., DEUTSCHE U., STORCK T., MUELLER-HILL B.;  
RL SUBMITTED (DEC-1989) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: X17483; G9843; -.  
FT NON\_TER 1  
FT NON\_TER 391  
SQ SEQUENCE 391 AA; 45309 MW; 1470DB94 CRC32;

Query Match 2.5%; Score 169; DB 5; Length 391;  
Best Local Similarity 37.3%; Pred. No. 1.14e-02;  
Matches 38; Conservative 29; Mismatches 35; Indels 0; Gaps 0;

Db N  
Dt 490 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 549  
Cp 1281 AATATCAGATGATCGAATCAATATATATATATATATATATATATATAT 1222  
Ct N I T C T H K \* N N S I I K L F \* Y I N  
Db N N N N N N N P S S S S N L I N  
Dt 550 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 591  
Cp 1221 AAGATCTTCAATTAATGACGATGACCTCCATTTTATCAAT 1180  
Ct K I F I N E R C T S N F I N

RESULT 15  
ID 081057 PRELIMINARY; PRT; 149 AA.  
AC 081057;  
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE KINETOCHORE (SKPLP)-LIKE PROTEIN.  
GN T18E12.16.  
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).

OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
OC CAPPALES; BRASSICACEAE; ARABIDOPSIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA ROUNSLEY S.D., LIN X., KAUL S., SHEA T.P., FUJII C.Y., MASON T.M.,  
RA SHEN M., RONNING C.M., FRASER C.M., SOMERVILLE C.R., VENTER J.C.;  
RT "Arabidopsis thaliana chromosome II BAC T18E12 genomic sequence";  
RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AC005313; G3548813; -.  
SQ SEQUENCE 149 AA; 17102 MW; C0311AAE CRC32;

Query Match 2.4%; Score 162; DB 10; Length 149;  
Best Local Similarity 36.0%; Pred. No. 6.80e-02;  
Matches 45; Conservative 26; Mismatches 54; Indels 0; Gaps 0;

Db C V V T E V P L O N V T G K I L S I V V  
Dt 107 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 166  
Cp 125 ACTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 184  
Ct Y V F T E V A L R N K G I K I L P F G A  
Db E Y C K K H V V D E E S D E F K T M D E  
Dt 167 TNGATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 226  
Cp 185 CTAATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 244  
Ct N N I K E D M V V I Q G N A F A S S H E

Db 227 ARGAR 231  
Dt 1:111:  
Cp 245 AAGAA 249  
Ct E

Search completed: Wed Nov 24 01:42:20 1999  
Job time: 1058 secs.











DT 01-JAN-1998 (TREMBLER, 05, CREATED)  
 DT 01-JAN-1998 (TREMBLER, 05, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLER, 08, LAST ANNOTATION UPDATE)  
 DE SEX DETERMINING PROTEIN.  
 CN SRY  
 OS MUS MUSCULUS MUSCULUS (EASTERN EUROPEAN MOUSE).  
 OC EUDAROTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA  
 OC SCUROGNATHI; MURIDAE; MURINAE; MUS.  
 [1]  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-B6JEL-YACH.  
 RX MEDLINE: 98043417.  
 RA ALBRECHT K.H., EICHER E.M.:  
 RT "DNA sequence analysis of Sry alleles (subgenus Mus) implicates  
 RT misregulation as the cause of C57BL/6J-t(P0S) sex reversal and  
 RT defines the SRY functional unit.".  
 RL GENETICS 147:1267-1277(1997).  
 DR EMBL: U70651; G2623367; -



Query Match 3.08; Score 204; DB 8; Length 394;  
Best Local Similarity 37.28; Pred. No. 7, 10e-07;  
Matches 73; Conservative 48; Mismatches 73; Indels 2; Gaps

Dbb S N C N I O I N K Y K S K Y L Q F R L I  
445 WSNAAVGGNAAYATHCRAATYAAATYAARLAWASNAATYATTCARITTYGANTNTH 504  
Dc :  
706 TGTCTCGCATCTTTTAAACCACTAATAATTAATLGAACAATCC - TTCATTTCG 649  
cp C H P N L \* I H S N K L E H L X F N F T

[illegible]



[illegible]

Dt	B87	GATTWMSGNGATNCARGNNAGTNTNCRBMGNARAAVWSWSWGARCNTTYTNTTY	946
Oy	893	CCTTTGGTGCGITTAACGTCGTT-CATAAACAACATAATGCAGAAATCAAGTT-ATTGT	950
Qt	F G G V K R R X Q * N Y N C K S S X L X		
Dd	E D Y A H H P S E I S C T L R A V R T A		
Dt	947	TNGRATTAAGCACCAYCCAMNSNGARATHTYMGNGCNTNGNMG	1006
Oy	951	-AGATATGATGACACCATCCAGAGAAATTAAGTGTACTACATTTGACACGACAGAAAGA	1009
Qt	D D Y A H H P R E I S A T I D T A R K K		
Dd	V G Q R R I L A I Y O P H R F S R L R E		
Dt	1007	CNGTNGRCAMGMNMGNAATHYTNCCNATHTAYHSCGCCALMGNTYSSNGNTNGMNG	1066
Oy	1010	AATAATCCCATTAAGAAGATTGTTCAGATTTTCAACACACACACTTCTCTGACACAG	1069
Qt	X P H K E V V A V F Q P H T F S R T Q A		
Dd	C I D S F P S A F K D A D E V L R T E V		
Dt	1067	ARTGAITHGAWSTYTTCNMNSCNCTTAYARGCAGCGAGATGYTNTNMGNCARG	1126
Oy	1070	CATTTTTAAGCATTTGCCAAGACTTANTGTAAAGCAGACGTCGTAATCTTAGTGAAA	1129
Qt	F L N E F A E S L C K A D R V F L C E I		
Dd	1127	TNT 1129 ..	
Dt	I I		
Oy	1130	TTT 1132	
Qt			
RESULT	3	PRELIMINARY; PRT; 960 AA.	
ID	Q25802;		
AC	Q25802;		
DT	01-NOV-1996 (TREMBL.REL. 01, CREATED)		
DT	01-NOV-1996 (TREMBL.REL. 01, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998 (TREMBL.REL. 08, LAST ANNOTATION UPDATE)		
DE	FRAMESHIFT.		
GN	RPOD.		
OS	PLASMIDIUM FALCIPARUM.		
OC	EUKARYOTA, ALVEOLATA, APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
KC	STRAIN-C10;		
KA	MEDLINE; 96346169.		
RA	WILSON R.J.M., DENNY P.W., PREISER P.R., RANGACHARI K., ROBERTS K.,		
RA	ROY A., WHITE A., STRATH M., MOORE D.J., MOORE P.W., WILLIAMSON D.H.;		
RT	"Complete gene map of the plasmid-like DNA of the malarial parasite		
RL	Plasmodium falciparum.";		
J. MOL. BIOL.	261:155-172(1996).		
DR	EMBL; X95273; E220245;..		
SQ	SEQUENCE 960 AA; 117987 MW; C9D612D2 CRC32;		
Query Match	3.1%; Score 210; DB 5; Length 960;		
Best Local Similarity	29.1%; Pred. No. 1,21e-07;		
Matches	141; Conservative 122; Mismatches 222; Indels 0; Gaps 0		
Dd	N F I A N S I S K N F K Y N L N N I I I		
Dt	1227	YAAATYTHFAVAAYSMNATHMSNAARAATTYAARTAVAYYTNAAVAATHAHAT	1286
CP	994	CAATGTGTGACTATTTCTCTGTGAGGTGTCGATATCATCATCACATATCATGATTTG	935
Qt	N C S T N F S W M V C I I Y N N L I C		
Dd	K Y L N N V I S K Y Y N Y S N I O L L I K		
Dt	1287	HAAATYATNAAYVAGYNATHAARTAVTAYAAVTAYAVSNAAVYTHCARNTYNATHAA	1346
CP	934	CAATGTGTGACTATTTCTCTGTGAGGTGTCGATATCATCATCACATATCATGATTTG	875
Qt	N C S F I E T T F N T T K R F * C F F N		
Dd	N I H N K W I L Y N I Y T Y L Y Y Y H		
Dt	1347	RAATYTHCAVAAYNATRGATHYNTLAAYATHTAYACNTAYATYTNATYTLATYACA	1406
CP	874	TATTTGTAACATCTAGCTTCTCTAATAACNACGACGAATTCAGCTAATGATTTAAAA	815
Qt	I C N I * L L I * I T N R N Y S * C I * N		



















RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE: 90326550.  
RA IKEDA M., WACHI M., JUNG H.K., ISHINO F., MATSUHASHI M.;  
RT "Nucleotide sequence involving *murG* and *murC* in the *mra* gene cluster  
RL region of *Escherichia coli*.";  
RN NUCLEIC ACIDS RES. 18:4014-4014(1990).  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE: 92334977.  
RA YUHA T., MORI H., NAGAI H., NAGATA T., ISHIMURA A., FUJITA N.,  
RA ISONO K., MIYOSHI K., NAKATA A.;  
RT "Systematic sequencing of the *Escherichia coli* genome: analysis of  
RL the 0-2.4 min region";  
RN NUCLEIC ACIDS RES. 20:3305-3308(1992).  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE: 97426617.  
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GODDEN M.A., ROSE D.J.,  
RA MAU B., SHAO Y.;  
RT "The complete genome sequence of *Escherichia coli* K-12";  
RL SCIENCE 277:1453-1474(1997).  
[4]  
RP MUTANT MURC3. SEQUENCE FROM N.A.  
RC STRAIN-CGSC 5988.  
RX MEDLINE: 97309380.  
RA EVELAND S.S., POMPELIANO D.L., ANDERSON M.S.;  
RT "Conditionally lethal *Escherichia coli* murein mutants contain point  
RL defects that map to regions conserved among murein and *folYl*  
RT poly-gamma-glutamate ligases: identification of a ligase  
RL superfamily.";  
RN BIOCHEMISTRY 36:6223-6229(1997).  
[5]  
RP SEQUENCE OF 1-14, AND CHARACTERIZATION.  
RC MEDLINE: 95324553.  
RA LIGER D., MASSON A., BLANOT D., VAN HEIJENOORT J., PARQUET C.;  
RT "Over-production, purification and properties of the  
RL uridine-diphosphate-N-acetylmuramate:L-alanine ligase from  
RT *Escherichia coli*.";  
RN EUR. J. BIOCHEM. 230:80-87(1995).  
CC -1- FUNCTION: CELL WALL FORMATION.  
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL-L-ALANINE -  
CC ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.  
CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
CC -1- SIMILARITY: BELONGS TO THE MURCEDE FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on how  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC  
DR EMBL: X52644; G42056; -;  
DR EMBL: X55034; G40859; -;  
DR EMBL: D10483; G216505; -;  
DR EMBL: AE000118; G1786279; -;  
DR EMBL: U67892; G2177094; -;  
DR PIR: J00545; CEECAM.  
DR PIR: S40601; S40601.  
DR ECOGENE: EG10619; MURC.  
DR PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;  
KM ATP-BINDING.  
KW NP BIND 126 132 ATP (POTENTIAL).  
FT MPTAGEN 344 344 G->: IN MURC3.  
QO SEQUENCE 491 AA; 53626 MW; C68E8B7D CRC32;

Query Match 94.8% Score 296; DB 1; Length 491;  
Best Local Similarity 42.5%; Pred No. 1,51e-21;  
Matches 79; Conservative 27; Mismatches 80; Indels 0; Gaps 0;

D D Y G H P T E V D A T I K A A R A  
Dc 1043 TNGTGAAGTAATGACGCATCAACCCAGTCGTGAGCAGCATTAACCGCNGMG 1102  
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Oy 947 TTGTGATGATTTATGCCAACCATTCACAAGAATAAGTGCTACATTGACACAGACCAA 1006  
V D D Y A H H P R E I S A T I D T A R K

G W P D K N L V M L F Q P H R F T R  
Db 1103 CNGGTGGCCAGAAATYTGNTGNATGYNCARCACCACAAVGGTTCACGCMG 1162  
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Oy 1007 AGAATATCCACATTAAGAATTTGTGAGATTTGCACACTCTTCTTAGACAC 1066  
K Y P H K E V V A V F Q P H T F S R T Q

D L Y D D F A N V L T Q V D T L L M L E  
Db 1163 GAGATYATGAGCAATYTCNNAAYGNTNYCACGCCACCAVGGTTCACGCMG 1222  
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Oy 1067 AAGCATTTTAATGAAATTTGCAAGAAATTTATGTAAGCAGATCGTATTTATGTG 1126  
A F L N E F A E S L C K A D R V F L C E

V  
Db 1223 ARGENT 1228  
| : | : |  
Oy 1127 AAATTT 1132  
I

RESULT 9  
ID MURC BUCAP STANDARD; PRF: 483 AA.  
Ac O51926;  
Dt 15-DEC-1998 (REL. 37, CREATED)  
Dt 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
Dt 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
De UDP-N-ACETYLMURAMATE--ALANINE LIASE (EC 6.3.2.8) (UDP-N-  
DE ACETYLMURAMOYL-L-ALANINE SYNTHETASE).  
MURC.  
GN BOCHNERA APHIDIOLA.  
OS BOCHNERA; PROTEOBACTERIA; GAMMA SUBDIVISION; BOCHNERA.  
OC [1]  
RX SEQUENCE FROM N.A.  
RP MEDLINE; 98087557.  
RA BAUMANN L., BAUMANN P.;  
RT "Characterization of itsz, the cell division gene of Buchnera  
aphidicola (endosymbiont of aphids) and detection of the product.";  
RL CURR. MICROBIOL. 36:85-89(1998).  
Cc -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).  
Cc -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL + L-ALANINE =  
Cc ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.  
Cc -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
Cc -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
Cc -1- SIMILARITY: BELONGS TO THE MURCODE FAMILY.

-----  
Cc This SWISS-PROT entry is copyright. It is produced through a collaboration  
Cc between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
Cc the European Bioinformatics Institute. There are no restrictions on its  
Cc use by non-profit institutions as long as its content is in no way  
Cc modified and this statement is not removed. Usage by and for commercial  
Cc entities requires a license agreement (See http://www.isb.ch/announce/  
Cc or send an email to license@isb-sib.ch).

-----  
DR EMBL; AF012886; G2738586; "  
KW PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;  
KM ATP-BINDING.  
FT NP\_BIND 124 130 ATP (POTENTIAL).  
SQ SEQUENCE 483 AA; 54852 MW; 23E8697B CRC32;

Best Match 3.9%; Score 266; DB 1; Length 483;  
Query Local Similarity 38.5%; Pred. No. 7.93e-17;  
Matches 67; Conservative 39; Mismatches 68; Indels 0; Gaps 0;

N D Y G H H P T E L S E T I K T I R K S



RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLOBER A., KELLEY J.M.,  
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T.T., HEDBLOW E., CORTON M.D.,  
RA UTTERRACK I.R., HANNA M.C., NGUYEN D.T., SAUDER D.M., BRANDON R.C.,  
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,  
RA GENHEIM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,  
RA VENTER J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
influenzae Rd.";  
RL SCIENCE 269:496-512(1995)  
CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMYL-L-ALANINE =  
CC ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMYL-L-ALANINE.  
CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: CYTOSOL; PLASMIC (PROBABLY).  
CC -1- SIMILARITY: BELONGS TO THE MRCDPEF FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on les-  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR DR EMBL; U32794; g1574695; -  
KW KW TIGR; H1139; -  
KW PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LGASE;  
KW ATP-BINDING.  
FT NP\_BIND 125 131 ATP (POTENTIAL).  
SO SEQUENCE 475 AA; 51994 MW; DE40C270 CRC32;  
  
Query Match 4.4%; Score 296; DB 1; Length 475;  
Best Local Similarity 39.8%; Pred. No. 1.51e-21;  
Matches 82; Conservative 38; Mismatches 85; Indels 1; Gaps 1;

DG N G K V R L V D D Y G H P T E V G V T  
DL 1009 AAYGGNARGNGNY-TNGNGAVGYATVAGNCNCAAGNGRGNGNGTAC 1067  
DY ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::  
QY 930 AATTGCACATAAGTATGTAGATGAATTAAGCACCATTCAGAAGATAGTGCTA 989  
QT N C K S R Y V D D Y A H H P R E I S A T  
DB I K A A R E G W G D K R I V M I F O P H  
DL 1068 NATHARCNCNNMGNGRGNGTGGGNGAVARRGNATHGMATGATHTTCARCNC 1127  
DY ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::  
QY 990 AATTGCACACACAGAAATATCCACAAAGAAAGTGTGCGAGTATTCAACAC 1049  
QT I D T A R K K Y P H K E V V A V F Q P H  
DB R Y S R T R D L F D D F V Q V L S Q V D  
DL 1128 YMGNTAVSMNGNMCNMGNGAVYTYTGYGAVYTYGTCNARGNYTNWSKARGTNGA 1187  
DY ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::  
QY 1050 CACTTTCCTGACACACAGCATTTTAAAGAAATTCACAAGTATGTAAGACG 1109  
QT T F S S R T Q A F L N E F A E S L C K A D  
DB A L I M L D V Y  
DL 1188 YGCANTTNATGATNGAVGINTIATG 1213  
DY ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::  
QY 1110 TCGTGTATCTTAGTGTGAATTTTGG 1135  
QT R V F L C E I F

RESULT 8  
ID ID MRCB\_ECOLI STANDARD: PR1: 491 AA.  
AC AC P17952; 007099;  
DT DT 01-NOV-1990 (REL. 16, CREATED)  
DT DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
DT DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE DE UDP-N-ACETYLMURAMATE--ALANINE LIASE (EC 6.3.2.8) (UDP-N-  
DN ACETYL-MURAMYL-L-ALANINE SYNTHETASE).  
MS MSCB.  
OS ESCHERICHA COLI.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
OC ESCHERICHTIA.  
LN [1]















[illegible]

QY	742	ACGCTTTGATGCTAGTGCATGCGTGGATTATGACGACTTCGCTCCGCAAAAT	801
QI		T A F D V Y V L D G E F Y D H F L S P Q Y	
Db		G D H T V L N A L A V I A I S Y L E K L	
Dc	781	GGNAYCAVACNGNTNNAAVCGNCTNGNATHGCAATHSMTAYVINGARARVTN	840
QY	802	GGTACCATCAAGTTTAAATGCATTAGCTGTAATTCGATTAATTCAGTATTGAGACGTA	861
QI		G D H T V L N A L A V I A I S Y L E K L	
Db		D V T N I K E A L E T F G G V K R R F N	
Dc	841	GAYTNACNAAYATHARAGRCNGNYNGARACNTTGGGNGGTNAARNGMNTTYAY	900
QY	862	GATGTTCAATATATAAGACGATTAGAAACGTTGGTGGCTTAACGTCGTTCAAT	921
QI		D V T N I K E A L E T F G G V K R R F N	
Db		E T T I A N Q V I V D D Y A H H P R E I	
Dc	901	GARCNACNATHGCAAGNATGATHGNGATGATYACCAVCAVCCMNGNARATH	960
QY	922	GAACATCAATTCGAATCAAGTTATTGTAGATGATTAGACACCATTCGCAAGAAATT	981
QI		E T T I A N Q V I V D D Y A H H P R E I	
Db		S A T I E T A R K K Y P H K E V V A V F	
Dc	961	WSMNCNACNATHGCAAGNATGATHGNGATGATYACCAVCAVCCMNGNARATH	1020
QY	982	AGTCTCAATTCGAATCAAGTTATTGTAGATGATTAGACACCATTCGCAAGAAATT	1041
QI		S A T I D T A R K K Y P H K E V V A V F	
Db		Q P H T T F S M R T Q A F L N A E R F A E S L S	
Dc	1021	CARCNACNATHGCAAGNATGATHGNGATGATYACCAVCAVCCMNGNARATH	1080
QY	1042	CACACACACCTTCTGTAGACAGCAAGATTTAAATGAATTTGACAGAAATTGAT	1101
QI		Q P H T T F S R T Q A F L N A E R F A E S L C	
Db		K A D R V F L C E I F G S I R E N T G A	
Dc	1081	AARCNACNAYMNGNTNYTNYTNGYGARATHHTTGCGMWSNATHMGCAARAAYACNGN	1140
QY	1102	AAACGACGTCGTGATTTATGTATGTAATTTGGCTCAATTAGAGAAATTTGCGCCA	1161
QI		K A D R V F L C E I F G S I R E N S G A	
Db		L T I O D L I D K I E G A S L I N E D S	
Dc	1141	YTNCTHTHARCAAYVYNNAATHGARGGCMWSNATHMGCAARAAYACNGN	1200
QY	1162	TTAACGATACAGTTTAAATGATTAATAATGAGGATGCTTATTAATGAATGATTT	1221
QI		L T I O D L I D K I G G A S F I N E D L	
Db		I N V L E O F D N A V V L F K G A G D I	
Dc	1201	ATHAAYGNTYNGARCAATRTTYGAYAAVCGNCTNGNTTYTTTAAAGGCGNCGNATYTH	1260
QY	1222	ATTATGATTTAGAACCAATTTGATGTAATGCTGTGTTTATTTATGCGGACGATGATTT	1281
QI		I N V L E Q F D N A V V L F M G A G D I	
Db		Q K L Q N A Y L D K L G M K N A F	
Dc	1261	CARAARTNCAARAAYGNTYTYNGNAAATRTYTGCAATGAATAAATGCTTT	1311
QY	1282	CAAAATTTACAAATGCAATTTTATGATTAATTAATGAGCATTAATAAATGCGCTTT	1332
QI		Q K L Q N A Y L D K L G M K N A F	
RESULT	2		
ID	MORC_BACSU	STANDARD;	PRT: 432 AA.
AC	P40778;		
DT	01-FEB-1995 (REL. 31, CREATED)		
DT	15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)		
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)		
DE	UDP-N-ACETYLUMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-		
DE	ACETYLUMURAMATE--L-ALANINE SYNTHETASE).		
GN	MORC.		
OS	BACILLUS SUBTILIS.		
OC	BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;		
OC	BACILLUS.		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 98048467.		



\*\*\*\*\*  
 WIDEWIDE (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MSPrch\_ntp n.a. - n.a. Smith-Waterman search, using a protein database  
 which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Nov 24 01:42:37 1999; Maspar time 151.81 Seconds  
 1509.426 Million cell updates/sec

Tubular output not generated.

Title: >US-09-103-287-1  
 Description: (1.1351) from US09103287.seq  
 Perfect Score: 6755  
 N.A. Sequence: 1 ATGAGTAGAGAGTTTATAT.....TTATATGTTTATATAGAG 1351  
 Comp: TACTATCTCCATAAATATA.....AATTATACAAATATATATCTC

Scoring table: TABLE bkttranslated2  
 Gap 30

Mmatch STD : Dbase 0; Query 0

Searched: 77977 segs, 84804879 bases x 2

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: swiss-prot37  
 1:swissprot

Statistics: Mean 78.980; Variance 158.854; scale 0.497

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	5272	78.0	437	1	MURC-STPAU	UDP-N-ACETYLUMURAMATE--	0.00e+00
2	2964	43.9	432	1	MURC-BAGSI	UDP-N-ACETYLUMURAMATE--	0.00e+00
3	310	4.6	433	1	MURC-PORGI	UDP-N-ACETYLUMURAMATE--	8.29e-24
4	309	4.6	505	1	MURC-SYNY3	UDP-N-ACETYLUMURAMATE--	1.21e-23
5	288	4.4	280	1	MURC-STNP7	UDP-N-ACETYLUMURAMATE--	7.24e-22
6	289	4.4	468	1	MURC-BORBU	UDP-N-ACETYLUMURAMATE--	5.00e-22
7	286	4.4	475	1	MURC-HAEIN	UDP-N-ACETYLUMURAMATE--	1.51e-21
8	296	4.4	491	1	MURC-ECOLI	UDP-N-ACETYLUMURAMATE--	1.51e-21
9	266	3.9	481	1	MURC-BUCAP	UDP-N-ACETYLUMURAMATE--	7.93e-17
10	252	3.7	481	1	MURC-TREPA	UDP-N-ACETYLUMURAMATE--	1.08e-14
11	183	2.7	441	1	VG34-HSV11	HYPOTHETICAL GENE 34 P	4.90e-05
12	180	2.7	449	1	MURC-HELPI	UDP-N-ACETYLUMURAMATE--	1.17e-04
13	173	2.6	396	1	RMAR-YEAST	MITOCHONDRIAL RIBOSOMA	8.57e-04
14	172	2.5	454	1	MURC-MYCTU	UDP-N-ACETYLUMURAMATE--	1.13e-03
15	167	2.5	758	1	CC27-YEAST	CELL DIVISION CONTROL	4.50e-03
16	160	2.4	395	1	SRY-MOUSE	SEX-DETERMINING REGION	2.94e-02
17	164	2.4	1024	1	RPOB-PLAFA	DNA-DIRECTED RNA POLYM	1.02e-02
18	162	2.4	1584	1	KYK1-DICDI	NON-RECEPTOR TYROSINE	1.73e-02
19	155	2.3	311	1	SRY-MUSSI	SEX-DETERMINING REGION	1.08e-01

C	20	152	2.3	432	1	NPX1-MOUSE	NEURONAL PENTRAXIN I P	2.31e-01
C	21	152	2.3	432	1	NPX1-RAT	NEURONAL PENTRAXIN I P	2.31e-01
C	22	152	2.3	537	1	ARP-PLAFA	ASPARAGINE-RICH PROTEI	2.31e-01
C	23	157	2.3	703	1	UVRD-MYCE	PUTATIVE DNA HELICASE	6.44e-02
C	24	152	2.3	797	1	PAFL-YEAST	TOPOISOMERASE II-ASSOC	2.31e-01
C	25	154	2.3	911	1	CA1B-BOVIN	COLLAGEN ALPHA 1(XI) C	1.39e-01
C	26	152	2.3	1024	1	SYIP-STPAU	ISOLEUCYL-TRNA SYNTHET	2.31e-01
C	27	157	2.3	1481	1	RPOD-ODOSI	DNA-DIRECTED RNA POLYM	6.44e-02
C	28	157	2.3	1682	1	MSP1-PLAF3	MEROZOITE SURFACE PROT	6.44e-02
C	29	147	2.2	225	1	VG2R-AMEPV	HYPOTHETICAL G2R PROTE	7.97e-01
C	30	147	2.2	337	1	R707-ACACA	MITOCHONDRIAL RIBOSOMA	7.97e-01
C	31	148	2.2	363	1	C1B-TRYBB	CYTOCHROME B (EC 1.10.	6.24e-01
C	32	147	2.2	430	1	NPX1-HUMAN	NEURONAL PENTRAXIN I P	7.97e-01
C	33	149	2.2	449	1	MURD-STPAU	UDP-N-ACETYLUMURAMYLAL	4.88e-01
C	34	150	2.2	502	1	LPLA-BACSU	LIPIDPROTEIN LPLA PRECU	3.81e-01
C	35	148	2.2	516	1	YJTI-ECOLI	HYPOTHETICAL 58.0 KD P	6.24e-01
C	36	149	2.2	575	1	RPOC-PLAFA	DNA-DIRECTED RNA POLYM	4.88e-01
C	37	150	2.2	708	1	GBF-DICDI	G-BOX BINDING FACTOR (	3.81e-01
C	38	146	2.2	791	1	LOM-CAMJE	ATP-DEPENDENT PROTEASE	1.02e+00
C	39	148	2.2	933	1	CC54-YEAST	CELL DIVISION CONTROL	6.24e-01
C	40	148	2.2	1905	1	TAGB-DICDI	PRESTARK-SPECIFIC PROT	6.24e-01
C	41	145	2.1	600	1	NIST-LACIA	NISTIN TRANSPORT ATP-BI	1.29e+00
C	42	144	2.1	981	1	YH48-YEAST	HYPOTHETICAL 113.2 KD	1.64e+00
C	43	145	2.1	1021	1	DPOM-NEUCR	PROBABLE DNA POLYMERAS	1.29e+00
C	44	145	2.1	1087	1	PER-DROVI	PERIOD CLOCK PROTEIN.	1.29e+00
C	45	144	2.1	1701	1	MSP1-PLAFM	MEROZOITE SURFACE PROT	1.64e+00

## ALIGNMENTS

RESULT 1  
 ID MURC-STPAU STANDARD: PRT: 437 AA.  
 AC 031211:  
 DT 15-UTL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
 DT 15-UTL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE UDP-N-ACETYLUMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-  
 ACETYLUMURAMYL-L-ALANINE SYNTHETASE).  
 GN MURC.  
 OS STAPHYLOCOCCUS AUREUS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;  
 NC STAPHYLOCOCCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA LOWE A.M., DERSEWICZ R.L.:  
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLUMURAMYL + L-ALANINE =  
 CC ADP + ORTHOPHOSPHATE + UDP-N-ACETYLUMURAMYL-L-ALANINE.  
 CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as the content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF034076; G2642659; -  
 CC PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;  
 CC ATP-BINDING.  
 CC KW ATP-BINDING.  
 CC FT NP\_BIND 108 114 ATP (POTENTIAL).  
 CC SO SEQUENCE 437 AA; 49176 MW; 64D6582 CRC32;  
 Query Match 78.0%; Score 5272; DB 1; Length 437;  
 Best Local Similarity 60.3%; Pred. No. 0.00e+00;  
 Matches 791; Conservative 319; Mismatches 201; Indels 0; Gaps 0;

Db M T H Y H F V G I K G S G M S S L A Q I  
 Dt 1 ATGACNCAVTAVCATYTGNGNATHAARGNWSGNATGWSMWSYVTGNCNCARATH 60



**THIS PAGE BLANK (USPTO)**







OY	902	GGTTAAACGTCGTTCAATGACATTCAAATTCGA-AATCAAGTTATTTGATGATTTAT	960
OT		V K R R F N K L Q L X N Q V I V D D Y	
Db	A H H P T E I G A T L K S A R		
Dc	961 GCNCAYCCACNACNGARATHGCGACNACVYTNAAFRSNCNGNA	1006	
OY	961 GCACACCATCCACAGAAATTAAGTCGTACCAATTCACACACACAGAA	1006	
OT	A H H P R E I S A T I D T A R		
RESULT	12		
ENTRY	A71699	#type complete	
TITLE	UDP-n-acetylmuramoylalanine--d-glutamate ligase (murD)	RP410	
ORGANISM	UDP-n-acetylmuramoylalanine--d-glutamate ligase (murD)	RP410	
DATE	21-Nov-1998	#sequence_revision 21-Nov-1998	#text_change
ACCESSIONS	A71699		
REFERENCE	A71630		
authors	Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Slicheritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.; Neebund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland, C.G.		
#journal	Nature (1998) 396:133-140		
#title	The genome sequence of Rickettsia prowazekii and the origin of mitochondria.		
#accession	A71699		
#status	preliminary; nucleic acid sequence not shown; translation not shown		
##molecule_type	DNA		
##residues	1-445	##label	AND
##cross-references	GB:AJ235271; GB:AJ235269; NID:93860788; PID:e1342711		
	PID:g3860967		
##experimental_source	strain Madrid E		
GENETICS			
#gene	murD; RP410		
SUMMARY	#length 445	#molecular_weight 50129	#checksum 5467
Query Match	2.8%; score 188; DB 2; Length 445;		
Best Local Similarity	38.4%; Pred. No. 2,58e-04;		
Matches	61; Conservative	31; Mismatches	67; Indels 0; Gaps 0;
Db	H E V V R I A H H F N I P I I S D I D L		
Dc	229 CAYGARGTNGTMMNATHGCMCAVCATTTAAAYATHCNCNTHATHTMSNGAYATGAYYTN	288	
OY	241 CATGAGAAATAGTACGTCGACATCATTCGAATTTAGATGTTGATTAATGATTTT	300	
OT	H E E I V R A H Q L K L D V V S Y N D F		
Db	F F E K S K N L K F I A I T G T N G K S		
Dc	289 TTTTGGARRARWNNARAAAYTNAATTTATHCNCNTHACNGACNACNAAFGNARWNN	348	
OY	301 TTAGACACGATTTATGATCATATACCTTCAGTTCGTACGTGACACATGCTAAACT	360	
OT	L G Q I I D Q Y T S V A V T G A H G K T		
Db	T T T A L I S H I L N S N		
Dc	349 ACNACNACNGCNYTNATHWSNCAATYTNAAAYSNAY	387	
OY	361 TCTACACACGTTATATATCATCATGTTATGATGCTGAT	399	
OT	S T T G L L S H V M N G D		
RESULT	13		
ENTRY	h36789	#type complete	
TITLE	hypothetical protein ORF34 - ictaluriid herpesvirus 1 (strain		
ORGANISM	aburn 1)		
#note	#formal_name ictaluriid herpesvirus 1		
DATE	17-Aug-1992	#sequence_revision 17-Aug-1992	#text_change
ACCESSIONS	A36789		
REFERENCE	A36804		
authors	Davison, A.J.		

```
#submission      submitted to GenBank January 1992
#description      Channel catfish virus: a new type of herpesvirus.
#accession        H36789
##residues        1-441
##molecule_type  DNA
#cross-references GB:M75136; NID:g3j31209; PID:g3j31244
REFERENCE
#authors          Davison, A.J
#journal           Virology (1992) 186:9-14
#title            Channel catfish virus: a new type of herpesvirus.
#cross-references MUID:92087490
#contents          annotation
#note             neither protein nor nucleic acid sequence is given
GENETICS
#gene             34..
SUMMARY
#length 441 #molecular-weight 50513 #checksum 3740

Query Match      2.7%: Score 183; DB 2; Length 441;
Best Local Similarity 42.7%; Pred. No. 9.19e-04;
Matches 44; Conservative 23; Mismatches 36; Indels 0; Gaps 0;

Dd   N W F C N M I A L F L E C F I N T P N K
Dt   765 YAATGGTATTGAAATGATGACAGCANTNTTYNNGARCTYTAHAATAACMCNAATAA 824
Cp   622 TAATACTTTTATACATTAATGTGCCATTTCTTGGAATCATCAAAAACATCATTAATAT 563
Cl   N T F . F N I M C H F L E C I K N I I N I

Dd   I G M I W M N E R Y Y L E N
Dt   825 RATHGNATGATHTGCATGAAGABRGNTATYYTNGARAAY 867
Cp   562 CTTGGAATATATCAGATGATGCAAAATCATATTTGTCTATAT 520
Cl   F E I I R M I E I N I C H N

RESULT 14
ENTRY    G64597 #type complete
TITLE     UDP-N-acetylmuramate-alanine ligase - Helicobacter pylori
          (strain 26695)
ORGANISM  #formal_name Helicobacter pylori
DATE       09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
          10-Oct-1997
ACCESSION G64597
REFERENCE  A64520
          Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
          Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
          H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
          J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
          Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
          McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
          Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
          Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
          Fujii, C.; Bowman, C.; Wathey, L.; Wallin, E.; Hayes,
          W.S.; Borodovsky, M.; Karpe, P.D.; Smith, H.O.; Fraser,
          C.M.; Venter, J.C.
          Nature (1997) 388:539-547
          The complete genome sequence of the gastric pathogen
          Helicobacter pylori.
#accession MUID:97394467
#status      preliminary; nucleic acid sequence not shown;
              translation not shown
##molecule_type DNA
##residues      1-449 ##label TOM
#cross-references GB:AEO00576; GB:AEO00511; NID:g2j313736; PID:g2j313743
              TIGR:HP0623
SUMMARY #length 449 #molecular-weight 50793 #checksum 3884

Query Match      2.7%: Score 180; DB 2; Length 449;
Best Local Similarity 50.8%; Pred. No. 1.95e-03;
Matches 32; Conservative 14; Mismatches 17; Indels 0; Gaps 0;
```







```
##molecule_type DNA
##residues 1-803 ##label ARN
##cross-references GB:AE001348; GB:AE001273; NID:g3329216; PID:g3329222
```

#journal Science (1998) 281:375-388  
#title Complete genome sequence of *Treponema pallidum*, the syphilis  
spirochete.







```

translocation not known
##molecule_type DNA
##residues 1-468 ##label KLE
##cross-references GB:AE001180; GB:AE000783; NID:g2668755; PID:g2668761
TIGR:BB0817

```

```

FUNCTION
#description
one, of the ligases responsible for the synthesis of
UDP-N-acetylmuramyl pentapeptide, an intermediate in
cell-wall biosynthesis
#pathway
peptidoglycan biosynthesis
CLASSIFICATION
#superfamily
UDP-N-acetylmuramate--alanine ligase
KEYWORDS
ATP; cell division; cell wall; ligase; P-loop; peptidoglycan
biosynthesis
FEATURE
#region
nucleotide-binding motif A (P-loop)
125-130
#length
475
#molecular-weight
51994
#checksum
7812
SUMMARY

```



Dc	1081	GCGCNGGAVTGCATTAAYTTTCGAATPHHTTYGGMCGMCMGCAAAVCGANGNAR	1140
Oy	1102	AAGACAGATCGTGATTCTTGATGTAAAATTTGGCTCATTAAGAAAATTTGGCGCA	1161
Dd	1141	YTNACMGTHGAGDLYTCCARGNNAARRTHCAIYAAGNMARYMATRGARGAGAYD	1200
Oy	1162	TTAACGATACAGATTATTAATGTATAAATTTGGAGGTGCATCGTCATTAAAGAATCTT	1221
Dd	1201	ACMNSGNNTYTTAAAGCACGACGAYAACBCGNTNTNATHHTTATTAAGGCGCGAGATH	1260
Oy	1222	ATTAACTTAATTAACCAATTTGATTAATCTGTGTGTTATTAATGAGCGACGGATATT	1281
Dc	1261	CABAART 1267	
Oy	1282	CANAART 1288	
Qt		Q K	
RESULT	2		
ENTRY	B70418	#type complete	
TITLE	UDP-N-acetylmuramate-alanine ligase - Aquifex aeolicus		
ORGANISM	#formal_name Aquifex aeolicus		
DATE	08-May-1998 #sequence_revision 08-May-1998 #text_change 21-Aug-1998		
ACCESSIONS	B70418		
REFERENCE	A70300		
*authors	Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Aubay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.		
#journal	Nature (1998) 392:353-358		
#title	The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.		
#cross-references	MUID:98196666		
#accession	B70418		
#status	preliminary; nucleic acid sequence not shown;		
	translation not shown		
##molecule_type DNA			
##residues	1-454 #label AOF		
##cross-references	GB:AE000736; NID:g2983763; PID:g2983764; GB:AE000652		
##experimental_source strain VES			
GENETICS			
#gene	murf		
CLASSIFICATION	#superfamily UDP-N-acetylmuramate--alanine ligase		
SUMMARY	#length 454 #molecular_weight 50893 #checksum 2346		
Query Match	4.6%; Score 308; DB 2; Length 454;		
Best Local Similarity	40.0%; Pred. No. 1,30e-19;		
Matches	114; Conservative 52; Mismatches 114; Indels 5; Gaps 5		
Dc	H N V Y N A L A T G V A L L E L G V S F		
Dd	831 HCAVAAVTNNIAAYAGCNYYNGCNCACNACG-GNGTNGCYTNGARTNGGNTNSNT	889	
Oy	807 CCATACATTTTAAATGATAGTAGCTATTTGCCATTTGTATTATTAAGACGACGATGT	866	
Qt	H T V L N A L A V I A L V I * R S * M L		
Dd	E V I K K S L E E F R N A E R R L E L K		
Dc	890 TYGARGTAAATTAARAARSNTNGRGRRTTYGMAAGCAGAMGMNGNTNGARTNA	949	
Oy	867 TACAATATTTAAAGAG-CATTAGAAACGTTGTGGTGGCTTAAACGTCGTTTC-AAAGAA	924	
Qt	Q I L K K x L E T F G G V K R R F x * N		
Dc	G Y Y K K N S P V Y D D Y G H H P T E I K		
Dd	950 ARGGTATTAARAARYSNCGNNTNGTATGATYAGCAVCACACNACNCRABATTA	1009	
Oy	925 ACTCAATTTGCAATCAAGTTAT-TGTA-GATGATTAATGACCACCACTCAAGAGAAATTA	982	
Qt	Y N C K S S Y x X D Y A C H C H P R E I S		

DB	1010	ARGCONGINATHAAVMSVYTMWGAAVATGTAAGCAACAAVAAAYATNTNNGNNGNTTC	1069
DE		AV I N S L R D M Y P D K N L L V V F Q	
OY	983	GTGCTACATTTGATCAACGACGACGAAGAAGATACATACATGAAGAGTTGTCGAGATTTC	1042
OT		A T I D T A C A R K K Y P H K E V V A V F Q	
DB	1070	ARCNCAVMGTAVMSNMGNACNTATATYNTNTTYARBAATYTG	1114
DE		P H R Y S R T Y Y L F E D F	
OY	1043	AACCAACACTTCTCTGAGAACCAAGACATTTTAAATGAATTTG	1087
OT		P H T F S R T Q A F L N E F	
RESULT	3		
ENTRY	S76722	#type complete	
TITLE	UDP-N-acetylmutaromycinolalanine--D-glutamate ligase (EC 6.3.2.9)		
ALTERNATE_NAMES	murC - Synecocystis sp. (strain PCC 6803)		
ORGANISM	protein sir1423		
ALTERNATE_NAMES	#formal name Synecocystis sp.		
DATE	PCC 6803		
REFERENCE	25-Apr-1997 #sequence-revision 25-Apr-1997 #text-change		
ACCESSIONS	21-Aug-1998		
REFERENCE	S76722		
#authors	S74322		
#journal	Kaneke, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, A.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Nario, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Matanabe, A.; Yamada, M.; Yasuda, M.; Tabara, S.		
#file	DNA Res. (1996) 3:109-136		
#title	Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.		
#cross-references	MUID:97061201		
#accession	S76722		
#status	nucleic acid sequence not shown; translation not shown		
#molecule-type	DNA		
#residues	1-505 #label KAN		
#cross-references	EMBL:D90916; GB:AB001339; NID:g1653715; PID:d1019367, P1D:g1653723		
#note	the nucleotide sequence was submitted to the EMBL Data Library, June 1996		
GENETICS:			
#gene	murC		
#start_codon	GTG		
KEYWORDS	ligase		
SUMMARY	#length 505 #molecular-weight 54513 #checksum 3888		
Query Match	4.6%; Score 309; DB 2; Length 505;		
Best Local Similarity	34.6%; Pred. No.9,45e-20;		
Matches	198; Conservative 91; Mismatches 277; Indels 6; Gaps 6;		
DB	617	ENGARATVYVYNGNGCANGATNGAGYAGWNSGAYGGMNSNTNACNARAYCAYCCNG	676
DE		E Y L V A E V D E S D G S L T K H H P E	
OY	452	GTGATTTATTTGCGTTTGAGCGATGGAATPAGACGTCATTTTAAGTTATAAACGTG	511
OT		D Y F A F E A C E Y R R H F L S Y K P D	
DB	677	ARAAGHNAHGNACAAATATGATYNGACAYCAYCAGCAGNACAYAVYMSNACNTNGCG	736
DE		I G I V T N I E L D H P D H Y S T L A E	
OY	512	ATACGAATATGACAAATATGATTTGATTCGATCCATCCGATATATTTCAAGATATATG	571
OT		Y A I M I N I D F D H P D Y F K D I N D	
DB	737	ARGNGNAAHHTYVYMGACTTGTG--RMSNCAVYTCGACACN--TNTHGGCTG	793
DE		V V E I F R T F E S H C Q T L I G C L	
OY	572	ATGTTTGTGATGCATCCAGAAATGCGCATTAATGTTAAAAAGCTTATTTATGCTTGCG	631
OT		V F D A F Q E M A C * C * K R Y I A W G	



[illegible]

DB	H V I O N A K R P T C M S F L I G D G N T G N O B G
Dc	361 CAYGNATHCARAAVYGCNARCCNACMSFTTYLTATHGNGATGNCNGNRBGN 420
Oy	382 CATGTATGAGTGGTGTAAGAAGACTCTCATTTTATGTATTGATGTCAGCAGATGAGGA 441
Qc	H V M N N G D K K K T S F L I G D G T G T G M G
Db	N E N S' E Y F V F E A C E Y R R H F L S
Dc	421 AAAGRAAAYNSNKRTAVTTYNTGYTAGRCVTGCAATATMGNAKNCATYYTYTNMSN 480
Oy	442 TTGCCGTAAAGTGTATTTTTCGTTTTGAGCGCATGGAATATAGACGTCCACTTTTAAGT 501
Qc	L P E S D Y F A F E A C E Y R R H F L S
Db	Y O P D Y A I M T N I D F D H P D Y F F S
Dc	481 TAICARCNCNMYTAGCNATHATGACAAATATGATTTGATCATCCNCAATATYYTYSN 540
Oy	502 TATAAACCTGTATTCGCAATATATGACAAATATGATTTGATGCATCCGATATTTCAA 561
Qc	Y K P D Y A I M T N I D F D H P D Y F F K
Db	S I D D V F D A F O E M A L O V N K G I
Dc	541 WSNATHGVAGYINTTYGAVGNTTYCARGARTGCVNTTCARSTNAAYARSGNTH 600
Oy	562 GATATATATGATTTGTTTGATGATCCAGAAGATGGCACATAAATGTTAAAAAGATTT 621
Qc	D I N D V F D A F O E M A H N V K K G I
Db	I A C G D D E H L P K I H A N V P V V Y
Dc	601 ATHGNTGYGSNGYVGYEARCAIYTMCNABARTHCAGANAAYTNCNONTNTNYAY 660
Oy	622 ATYGCTGGGTGATGATGAACAATCTCCGRAATTAAGACAGATGTCCATTTATATAC 681
Qc	I A W G D E D E F Q A R I E A D V P T I E G
Db	Y G T G E E N D F Q A R I E I V K S T I E G
Dc	661 TAYGONAANGSNGRGRARAAYGYTYCYCARGCMNGAAYATHGTNAARSNBCNARBGN 720
Oy	682 TATGATTTAAAGTTCGGATTCGATTTATGATTCGCAAAATTTCAATTCGATTAAGT 741
Qc	Y G F K D S D D I Y A Q N I Q I T D K G
Db	T T F D V F V R N T F Y D T F Y I P A Y
Dc	721 ACNACNTTYGAGYNTTYGGMNMAAYACNTTYTAGAYACNTTYAATVATHCCNCNTAY 780
Oy	742 ACTGTTTGTAGTCGTAAGTGATGGATGATTTATGATCACTCTCGTCGACAAATAT 801
Qc	T A F D V V D G E F Y D H F L S P Q I
Db	G H H N V L N S L A V I A L C H Y S E I
Dc	781 GGNCAVCYAAYGNTYNMAAYWSNYTNGCNTNATHGCAVYNTGYCAVAYCARBARATH 840
Oy	802 GGTCGACATACGTTTAANTGATTAAGCGTATTAAGCATTAAGTATTTNAGAGCTA 861
Qc	G D H T V L N A L A V I A I S Y L E A K L
Db	D S S I I K H A L K S F G G V K R R F N
Dc	841 GAYSMSNATHAHARCAVYGCNTYNAARSWTITTYGNGNGTNAARSGNNTYAY 900
Oy	862 GATGTACAAATATTAAGAAGCATTAAGAAGCTTTGGTGCTTAAGAGTGTTCAT 921
Qc	D V T N I K E A L E T F G G V K R F N
Db	E K Q L G D Q V L I D D Y A A H H P T E I
Dc	901 GARABCARBYTNGSNGRGNTYNTATHGAYTAYGACMACVACAYCANCBNCRARTH 960
Oy	922 GAACCTACAAATTCGAATTCATTTATGTATGATTAAGCATTAAGCATTAAGCAAGAAAT 981
Qc	E T T I A N Q V I V D D Y A H C H P R E I
Db	K V T I E A A R Q K Y P D R E I V A V F
Dc	961 AARGNACNATHGARGCNGMGCARRAATATVACGAYMGANGARHTHNCNCTNTNY 1020
Oy	982 AGTGCTACAAATTCGACAGCAGCAAGAAATATCCACATTAAGAAAGTGTTCAGATTT 1041
Qc	S A T T I D T A R K K Y P H K E V A V F
Db	Q P H T F T R T Q O F L D D E F A E S L S
Dc	1021 CARCNCNACNTTYACNMNACNACARCATTYTGDAGARTTYGCNGBARSNTNSN 1080
Oy	1042 CAACCACACACTTTCCTAACAACAGCATTTTNAAGATTTTGCAGAAAGSLTTC 1101
Qc	Q P H T F S R T Q A F L N E F A E A S L T C
Db	G A D C V Y L C D I F G S A R E N A G K



\*\*\*\*\*  
 WISEN (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MSrch\_ntp n.a. - n.a. Smith-Waterman search, using a protein database  
 which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Nov 24 01:12:22 1999; Maspar time 218.82 Seconds  
 1484.328 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-1  
 Description: (1.1351) from US09103287.seq  
 Perfect Score: 6755  
 N.A. Sequence: 1 ATGAGTAAGAGTTTATAT.....TTAATATGTTTAAATAGAG 1351  
 Comp: TACTCATTCCTCAAAATATA.....AATATACAAATATATCTC

Scoring table: TABLE bkttranslated2  
 Gap 30

Nmatch STD : Dbase 0; Query 0

Searched: 122810 seqs, 120205779 bases x 2

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database:

pir60  
 1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 77.827; Variance 184.008; scale 0.423

Prod. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	2964	43.9	432	2	C69662	UDP-N-acetylmutamate-	0.00e+00
2	308	4.6	454	2	B70418	UDP-N-acetylmutamate-	1.30e-19
3	309	4.6	505	2	S76722	UDP-N-acetylmutamoyla	9.45e-20
4	289	4.4	468	2	H70201	UDP-N-acetylmutamate-	2.30e-18
5	266	4.4	475	2	E64185	UDP-N-acetylmutamate-	5.94e-18
6	266	4.4	491	1	CEECAM	UDP-N-acetylmutamate-	5.94e-18
7	262	3.8	495	2	C71679	UDP-n-acetylmutamate-	2.22e-13
8	259	3.8	803	2	A71475	Probable muramate-Ala	5.50e-13
9	252	3.7	481	2	C71338	Probable UDP-N-acetyl	4.48e-12
10	204	3.0	763	2	S17998	gene COX1 intron 4 pr	3.83e-06
11	188	2.9	449	2	B71917	udp-n-acetylmutamate-	1.90e-05
12	188	2.8	445	2	A71699	UDP-n-acetylmutamoyla	2.58e-04
13	183	2.7	441	2	H36789	hypothetical protein	9.19e-04
14	180	2.7	449	2	G64597	UDP-N-acetylmutamate-	1.95e-03
15	177	2.6	234	2	S14469	asparagine-rich prote	4.10e-03
16	177	2.6	307	2	A45600	asparagine-rich blood	4.10e-03
17	173	2.6	368	1	R381M1	ribosomal protein var	1.09e-02
18	170	2.5	369	1	B70220	conserved hypothetical	2.25e-02
19	169	2.5	391	2	S14577	asparagine-rich prote	2.86e-02

C	20	172	2.5	494	2	D70579	probable murC protein	1.39e-02
	21	167	2.5	758	2	S45825	cell division control	4.59e-02
	22	162	2.4	149	2	T02709	kinetochore SKP1 prot	1.48e-01
	23	160	2.4	251	1	HBSHBC	hemoglobin beta (CNA)	2.34e-01
	24	160	2.4	386	2	S58755	ribosomal protein VAR	2.34e-01
	25	160	2.4	395	2	S43344	sex-determining prote	2.34e-01
	26	160	2.4	395	2	S35565	sex-determining prote	2.34e-01
	27	164	2.4	1024	1	RN20BF	DNA-directed RNA poly	9.28e-02
	28	165	2.4	1182	2	G71607	Probable integral mem	7.35e-02
	29	161	2.4	1802	2	G71616	hypothetical protein	1.86e-01
	30	157	2.3	99	2	S44632	f2b7.3 protein - Cae	4.62e-01
	31	157	2.3	251	1	HBSHBC	hemoglobin beta (CNA)	4.62e-01
	32	153	2.3	407	2	C45600	asparagine-rich blood	1.14e+00
	33	152	2.3	412	2	S23208	mRNA maturase b12 - y	1.40e+00
	34	155	2.3	415	2	B71614	probable multiple tra	7.22e-01
	35	158	2.3	516	2	S34525	hypothetical protein	3.68e-01
	36	152	2.3	537	2	A23770	asparagine-rich prote	1.40e+00
	37	153	2.3	637	2	S78171	heme lyase yefJ - Rec	1.12e+00
	38	157	2.3	703	2	I64226	DNA helicase II (mutb	4.62e-01
	39	152	2.3	797	2	S53590	hypothetical protein	1.40e+00
	40	154	2.3	1024	2	S18251	collagen alpha 1(XI)	9.02e-01
	41	157	2.3	1060	2	S06286	major merocline surfa	4.62e-01
	42	155	2.3	1256	2	S14556	asparagine-rich prote	7.22e-01
	43	157	2.3	1481	2	S78373	DNA-directed RNA poly	4.62e-01
	44	152	2.3	1817	2	D71606	hypothetical protein	1.40e+00
	45	151	2.2	1021	2	I40805	collagenase - Clostri	1.74e+00

## ALIGNMENTS

RESULT 1  
 ENTRY  
 TITLE  
 ORGANISM  
 DATE  
 ACCESSIONS  
 REFERENCE  
 #authors

C69662 #type complete  
 UDP-N-acetylmutamate-alanine ligase (EC 6.3.2.8) murC -  
 Bacillus subtilis  
 #formal\_name Bacillus subtilis  
 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change  
 17-Mar-1999  
 C69662; S71002  
 A65580  
 Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;  
 Allion, G.; Azavedo, V.; Bertero, M.G.; Bessières, P.;  
 Bojotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,  
 A.; Braun, M.; Briganti, S.C.; Bron, S.; Brouillet, S.;  
 Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.;  
 Choi, S.K.; Codani, J.J.; Conneron, I.F.; Cummings, N.J.;  
 Daniel, R.A.; Denizot, F.; Devine, K.M.; Desterhoef, A.;  
 Ehlich, S.D.; Emerson, P.T.; Ehtian, K.D.; Errington, J.;  
 Fabret, C.; Ferrari, E.; Foulger, D.; Fultz, C.; Fujita,  
 M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Gilm,  
 S.Y.; Glaeser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;  
 Giuseppe, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,  
 C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.;  
 Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;  
 Kaasahara, Y.; Klaer-Blanchard, M.; Klein, C.; Kobayashi,  
 Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.;  
 Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;  
 Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;  
 Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,  
 M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,  
 M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,  
 V.; Pohl, T.M.; Portetlelle, D.; Porrolik, S.; Prescott,  
 A.M.; Presecan, E.; Pulic, P.; Purnelle, B.; Rapoport, G.;  
 Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;  
 Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.;  
 Schleich, S.; Schroeter, R.; Scottio, F.; Sekiguchi, J.;  
 Sekowska, A.; Seror, S.J.; Serro, P.; Shin, B.S.; Solido,  
 B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;  
 Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;  
 Tepser, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;  
 Vandenbol, M.; Vannier, F.; Vassart, A.; Viart, A.;  
 Wandut, R.; Wedler, E.; Wedler, H.; Welltzeneger, T.;  
 Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,  
 K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;



**THIS PAGE BLANK (USPTO)**



CC NUMBER OF SEQUENCES: 7  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: NIXON & VANDERHAYE P.C.  
 CC STREET: 1100 NORTH GLEBE ROAD  
 CC CITY: ARLINGTON  
 CC STATE: VIRGINIA  
 CC COUNTRY: U.S.A.  
 CC ZIP: 22201-4714  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patentin Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/273,846  
 CC FILING DATE: 12-JUL-1994  
 CC CLASSIFICATION: 530  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: WILSON, MARY J.  
 CC REGISTRATION NUMBER: 32,955  
 CC REFERENCE/DOCKET NUMBER: 555-29  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (703) 816-4000  
 CC TELEFAX: (703) 816-4100  
 CC INFORMATION FOR SEQ ID NO: 1:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 29 base pairs  
 CC TYPE: nucleic acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: DNA (genomic)  
 CC SEQUENCE 290 BP; 17 A; 34 C; 8 G; 11 T; 220 OTHER.

Query Match 1.7%; Score 23; DB 1; Length 242;  
 Best Local Similarity 12.6%; Pred. No. 7,69e-01;  
 Matches 14; Conservative 28; Mismatches 68; Indels 1; Gaps 1;  
 Db 53 DUREPDESCRFPROC\*FXPCALLEDFROMCFRFORM\*FXPCALLEDFROMCFRCONV 112  
 Cp 1166 GTTAATGCCAGAAATTTCTTAATGAGCCAAA-ATTTCATAGAAATGACGATC 1108  
 Db 113 T\*FXPCALLEDFROMON\*\*\*CALLEDFROMPROCEDURESETPATCCRA 163  
 Cp 1107 TCGTTTACATTAACCTTCTGCAAAATTCATTAAATAATGCTGTGTTCTAGA 1057

RESULT 14  
 ID 5189151-1 STANDARD; DNA; UNC: 841 BP.  
 AC xxxxxx  
 DT 01-JAN-1900  
 DE Patent No. 5189151.  
 CC Patent No. 5189151  
 CC APPLICANT: BAUDRY, BERNADETTE;LEVINE, MYRON M.  
 CC TITLE OF INVENTION: HIGHLY SPECIFIC DNA PROBE FOR  
 CC ENTEROAGREGATIVE ESCHERICHIA COLI  
 CC NUMBER OF SEQUENCES: 1  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/07/524,100  
 CC FILING DATE: 16-MAY-1990  
 CC SEQ ID NO:1:  
 CC LENGTH: 776  
 CC Sequence 841 BP; 296 A; 109 C; 128 G; 243 T; 65 other;

Query Match 1.7%; Score 23; DB 5; Length 776;  
 Best Local Similarity 69.5%; Pred. No. 7,69e-01;  
 Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
 Db 296 CTTATGAACCAAAATGCAATTAATGATGATTTGTAATTTAAATATATT 354  
 Oy 509 CTGATTACCAATTAATGCAAAATTAATGATTTGATCATCTGATTTATTTCAAGATATT 567

RESULT 15

ID US-08-388-672A-22 STANDARD; DNA; UNC: 965 BP.  
 AC xxxxxx  
 DT Sequence 22, Application US/08388672A  
 DE Sequence 22, Application US/08388672A  
 CC Patent No. 5795961  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Wallace, T. Paul  
 CC APPLICANT: Harris, William J.  
 CC APPLICANT: Carr, Frank J.  
 CC APPLICANT: Old, Lloyd J.  
 CC APPLICANT: Welt, Sydney  
 CC APPLICANT: Kitamura, Kunio  
 CC TITLE OF INVENTION: Recombinant Human Anti-Lewis B  
 CC NUMBER OF SEQUENCES: 25  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Felte and Lynch  
 CC STREET: 805 Third Avenue  
 CC CITY: New York  
 CC STATE: New York  
 CC COUNTRY: U.S.A.  
 CC ZIP: 10022  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patentin Release #1.0, Version #1.30  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/388,672A  
 CC FILING DATE: 14-FEB-1995  
 CC CLASSIFICATION:  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Hanson, No. 5795961man D.  
 CC REGISTRATION NUMBER: 30,946  
 CC REFERENCE/DOCKET NUMBER: LUD 5409  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: 212-688-9200  
 CC TELEFAX: 212-838-3884  
 CC INFORMATION FOR SEQ ID NO: 22:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 965 base pairs  
 CC TYPE: nucleic acid  
 CC STRANDEDNESS: unknown  
 CC TOPOLOGY: unknown  
 CC MOLECULE TYPE: DNA (genomic)  
 CC SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

Query Match 1.7%; Score 23; DB 3; Length 965;  
 Best Local Similarity 13.2%; Pred. No. 7,69e-01;  
 Matches 10; Conservative 38; Mismatches 28; Indels 0; Gaps 0;  
 Db 854 YVCVGRSVDGSGGYTIVTSVHVKKDMTSSSSASVGRVTTCSRSTHGNNTY 913  
 Cp 305 CCTAAAAATCATTTATTAACACATCTAATTCATGATGACGATCAATTTCT 246  
 Db 914 WYKGAAYVRSNGV 929  
 Cp 245 TCATGCTACTCGCA 230

Search completed: Sat Nov 27 11:51:35 1999  
 Job time : 104 secs.



CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 965 base pairs
CC	TYPE: nucleic acid
CC	STRANDEDNESS: unknown
CC	TOPOLOGY: unknown
CC	MOLECULE TYPE: DNA (genomic)
SQ	SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.
	Query Match            1.8%; Score 24; DB 3; Length 965;
	Best Local Similarity 19.8%; Pred No. 2,46e-01;
	Matches       23; Conservative     48; Mismatches 44; Indels    1; Gaps    1
Db	783 RHUVHVSQGVRSSTCTCTASDYYTTSYWGAVNRGRMGDIYGGYTNNGKRGV-TAADTSSN 841
	: ::::: : :   :
Oy	655 ATTGACAGCAGATGTCTCCATTATTACTATTACTATTCGATTTAAAGCATTCGGATGACATTATTCCT 714
Db	842 SRSSTAYADTVVYVCVRGRSDSCGDGWGTFYTVVSHUYKDMTSSSAAGVDRV 897
	: :
Oy	715 CAATAATATTCGAATTAGCGATAAAGCTCTGCTTTTGATGTGTATGTCGATGCAGTGCA 770

[illegible]

Db 65 NNVNNNNA 73  
          : : :  
Cp 737 TTATCCGTA 729

```

RESULT 12
ID PCT-US95-11934-97 STANDARD; DNA; UNC; 82 BP.
AC xxxxxx
DT
DE Sequence 97, Application PC/TUS9511934
CC Sequence 97, Application PC/TUS9511934
CC GENERAL INFORMATION:
CC APPLICANT: Cytogen Corporation
CC TITLE OF INVENTION: Antigen Binding Peptides (abtlies) From
CC TITLE OF INVENTION: Peptide Libraries
CC NUMBER OF SEQUENCES: 103
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Pennie & Edmonds
CC STREET: 1155 Avenue of the Americas
CC CITY: New York
CC STATE: New York
CC COUNTRY: USA
CC ZIP: 10036
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DO5
CC SOFTWARE: Patentln Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/11934
CC FILING DATE: 20-SEP-1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Mistrock, S. Leslie
CC REGISTRATION NUMBER: 18,872
CC REFERENCE/DOCKET NUMBER: 1101-196-228
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 790-9090
CC TELEFAX: (212) 869-9741/88864
CC TELEX: 66141 PENNIE
CC INFORMATION FOR SEQ ID NO: 97:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 82 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA (genomic)
CC SEQUENCE 82 BP; 1 A; 2 C; 10 G; 8 T; 61 OTHER.

Query Match 1.7%; Score 23; DB 4; Length 82;
Best Local Similarity 13.2%; Pred. NO. 7.69e-01;
Matches 10; Conservative 19; Mismatches 47; Indels 0; Gaps 0;

Db 2 TGTGTCGCAcNNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNB 61
CP 999 TGTGTCAATTGTACCACTAATTCTCTGTGATGCTGTGCATATCATCTACAACTG 940
Db 62 NBNBNNBNNBGGTT 77
CP 939 ATTGCAATTGTACTT 924

RESULT 13
ID US-08-273-846-1 STANDARD; DNA; UNC; 290 BP.
AC xxxxxx
DT
DE Sequence 1, Application US/08273846
CC Sequence 1, Application US/08273846
CC Patent No. 5641863
CC GENERAL INFORMATION:
CC APPLICANT: SCHREIBER, ALAN D.
CC APPLICANT: PARK, JONG-GU
CC TITLE OF INVENTION: METHODS OF STIMULATING PHAGOCYTOSIS

```



CC ZIP: 19406-0939  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette  
CC COMPUTER: IBM Compatible  
CC OPERATING SYSTEM: DOS  
CC SOFTWARE: FASTSEQ Version 1.5  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/878,106  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/357,962  
CC FILING DATE: 16-DEC-1994  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Jervis, Herbert H  
CC REGISTRATION NUMBER: 31,171  
CC REFERENCE/DOCKET NUMBER: P50278  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 610-270-5019  
CC TELEFAX: 610-270-5090  
CC TELEX:  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 108 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: Genomic DNA  
CC HYPOTHETICAL: NO  
CC ANTI-SENSE: NO  
CC FRAGMENT TYPE:  
CC ORIGINAL SOURCE:  
CC SEQUENCE 108 BP: 20 A; 23 C; 12 G; 10 T; 43 OTHER.  
SQ  
Query Match 1.8%; Score 24; DB 3; Length 108;  
Best Local Similarity 31.6%; Pred. No. 2.46e-01;  
Matches 18; Conservative 22; Mismatches 17; Indels 0; Gaps 0;  
DB 39 RTCNTWYCHACHACHACHCHRTDAGCGMCCMRGMRGACGAYWCHGDTMT 95  
CP 1278 ATCACCCTGCACCCATTAATAAACACAGCATTTCAATGTCTTAATACATTAAAT 1222  
RESULT 9  
ID US-08-357-962-2 STANDARD; DNA; UNC; 108 BP.  
AC xxxxxx  
DT  
DE Sequence 2, Application US/08357962  
CC Sequence 2, Application US/08357962  
CC Patent No. 5668263  
CC GENERAL INFORMATION:  
CC APPLICANT: Hoyer, Lois  
CC APPLICANT: Livi, George  
CC APPLICANT: Shatzman, Allan  
CC TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES  
CC NUMBER OF SEQUENCES: 9  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: SmithKline Beecham Corporation  
CC STREET: 709 Swedeland Road  
CC CITY: King of Prussia  
CC STATE: PA  
CC COUNTRY: USA  
CC ZIP: 19406-0939  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette  
CC COMPUTER: IBM Compatible  
CC OPERATING SYSTEM: DOS  
CC SOFTWARE: FASTSEQ Version 1.5  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/357,962  
CC FILING DATE: 16-DEC-1994  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER:  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Jervis, Herbert H  
CC REGISTRATION NUMBER: 31,171  
CC REFERENCE/DOCKET NUMBER: P50278  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 610-270-5019  
CC TELEFAX: 610-270-5090  
CC TELEX:  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 108 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: Genomic DNA  
CC HYPOTHETICAL: NO  
CC ANTI-SENSE: NO  
CC FRAGMENT TYPE:  
CC ORIGINAL SOURCE:  
CC SEQUENCE 108 BP: 20 A; 23 C; 12 G; 10 T; 43 OTHER.  
SQ  
Query Match 1.8%; Score 24; DB 2; Length 108;  
Best Local Similarity 31.6%; Pred. No. 2.46e-01;  
Matches 18; Conservative 22; Mismatches 17; Indels 0; Gaps 0;  
DB 39 RTCNTWYCHACHACHACHCHRTDAGCGMCCMRGMRGACGAYWCHGDTMT 95  
CP 1278 ATCACCCTGCACCCATTAATAAACACAGCATTTCAATGTCTTAATACATTAAAT 1222  
RESULT 10  
ID US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.  
AC xxxxxx  
DT  
DE Sequence 22, Application US/08388672A  
CC Sequence 22, Application US/08388672A  
CC Patent No. 5795961  
CC GENERAL INFORMATION:  
CC APPLICANT: Wallace, T. Paul  
CC APPLICANT: Harris, William J.  
CC APPLICANT: Carr, Frank J.  
CC APPLICANT: Old, Lloyd J.  
CC APPLICANT: Welt, Sydney  
CC APPLICANT: Kitamura, Kunio  
CC TITLE OF INVENTION: Recombinant Human Anti-Lewis B  
CC NUMBER OF SEQUENCES: 25  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Feltz and Lynch  
CC STREET: 805 Third Avenue  
CC CITY: New York  
CC STATE: New York  
CC COUNTRY: U.S.A.  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/388,672A  
CC FILING DATE: 14-FEB-1995  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Hanson, No. 5795961man D.  
CC REGISTRATION NUMBER: 30,946  
CC REFERENCE/DOCKET NUMBER: LUD 5409  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 212-688-9200  
CC TELEFAX: 212-838-3884  
CC INFORMATION FOR SEQ ID NO: 22:







```

CC NAME: KOKULIS, Paul N.
CC REGISTRATION NUMBER: 16773
CC REFERENCE/DOCKET NUMBER: PNK/5544/202253/DJP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-861-3000
CC TELEFAX: 202-822-0944
CC TELEX: 6714627 CUSH
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1200 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA (genomic)
CC SEQUENCE 1200 BP: 406 A; 224 C; 230 G; 323 T; 17 OTHER.

SO
    Query Match          2.0%; Score 27; DB 2; Length 1200;
    Best Local Similarity 46.2%; Pred. No. 7,166-03;
    Matches 18; Conservative 15; Mismatches 6; Indels 0; Gaps 0;

Db
    324 GACACATCAATYHHHATYCCYHAHAHVCCAAAACG 362
    |||||:||||:||||:||||:||||:|||||
Cp
    781 GATCATAAACTCACATCCACATACATCAAAAGCG 743

RESULT      4
ID US-08-357-264-2 STANDARD; DNA; UNC; 1200 BP.
AC xxxxxx
DT
DE Sequence 2, Application US/08357264
CC Sequence 2, Application US/08357264
CC Patent No. 5541077
CC GENERAL INFORMATION:
CC APPLICANT: BORNIE Mr., James P.
CC APPLICANT: MATTHEWS MS., Ruth C.
CC TITLE OF INVENTION: FUNGAL STRESS PROTEINS
CC NUMBER OF SEQUENCES: 10
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Cushman, Darby & Cushman
CC STREET: 1100 New York Avenue, N.W.
CC CITY: Washington
CC STATE: DC
CC COUNTRY: USA
CC ZIP: 20005-3918
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/357,264
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/152669
CC FILING DATE: 16-NOV-1993
CC APPLICATION NUMBER: US 663897
CC FILING DATE: 14-MAR-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: KOKULIS, Paul N.
CC REGISTRATION NUMBER: 16773
CC REFERENCE/DOCKET NUMBER: PNK/5544/202253/DJP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-861-3000
CC TELEFAX: 202-822-0944
CC TELEX: 6714627 CUSH
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1200 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA (genomic)

```

```

SQ SEQUENCE 1200 BP; 406 A; 224 C; 230 G; 323 T; 17 OTHER.

Query Match          2 0%; Score 27; DB 1; Length 1200;
Best Local Similarity 46.2%; Pred. No. 7.1e-03;
Matches      18; Conservative    15; Mismatches     6; Indels     0; Gaps     0

Db       324 GAACAYCAAAYYHHHYAACYCAHHAHVHCYCAAACAG 362
        |||||:|||||:::||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Cc       781 GATCATAAACTCACATCCATCACAATACATCAAAAAGCAG 743

RESULT              5
ID US-08-357-962-2 STANDARD; DNA: UNC; 108 BP.
AC xxxxxx

DE Sequence 2, Application US/08357962
CC Sequence 2, Application US/08357962
CC Patent No. 5668263
CC GENERAL INFORMATION:
CC APPLICANT: Hoyer, Lois
CC APPLICANT: Livai, George
CC APPLICANT: Shatzman, Allan
CC TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406-0939
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq Version 1.5
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/357,962
CC FILING DATE: 16-DEC-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jervis, Herbert H
CC REGISTRATION NUMBER: 31,171
CC REFERENCE/DOCKET NUMBER: P50278
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5019
CC TELEFAX: 610-270-5090
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 108 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: Genomic DNA
CC HYPOTHEICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE:
CC ORIGINAL SOURCE:
CC SEQUENCE 108 BP; 20 A; 23 C; 12 G; 10 T; 43 OTHER.

Query Match          1.9%; Score 25; DB 2; Length 108;
Best Local Similarity 32.6%; Pred. No. 7.7e-02;
Matches      14; Conservative    20; Mismatches     9; Indels     0; Gaps     0;

Db       48 HACHACHACHCHRTDACYGWCWCCMRGRGHAGYGAYCH 90
        ::||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Cc       324 TACTTCAGTAGCTGTACTGTGTGCACATGGTAATAACTTTTACA 366
```







WIPSE RELEASE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPerch.un n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 11:49:51 1999; MasPar time 99.13 Seconds

Tabular output not generated. 1178.682 Million cell updates/sec

Title: >US-09-103-287-1  
(1-1351) from US09103287.seq

Description: 1 ATCAGTAAAGAGTTTATAT.....TTATATGTTTATATAGAG 1351  
Perfect Score: 1351  
N.A. Sequence: 1 ATCAGTAAAGAGTTTATAT.....TTATATGTTTATATAGAG 1351  
Comp: TACTCATCTCTCAAAATATATA.....AATATCAAAATATATCTC

Scoring table: TABLE default

Gap 6

Mmatch STD: Dbase 0; Query 0

Searched: 165359 segs, 43243793 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: n-Issued  
1:5A\_COMB 2:5B\_COMB 3:5C\_COMB 4:PCT9\_COMB 5:backfiles1

Statistics: Mean 8.554; Variance 5.055; scale 1.692

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Query Length	ID	Description	Pred. No.
C 1	59	4.4	7218	2	US-08-232- Sequence 14, Applicati	6.73e-22
C 2	27	2.0	215	1	US-08-238- Sequence 5, Applicati	7.16e-03
C 3	27	2.0	1200	1	US-08-672- Sequence 2, Applicati	7.16e-03
C 4	27	2.0	1200	1	US-08-357- Sequence 2, Applicati	7.16e-03
C 5	25	1.9	108	2	US-08-357- Sequence 2, Applicati	7.72e-02
C 6	25	1.9	108	3	US-08-878- Sequence 2, Applicati	7.72e-02
C 7	26	1.9	215	1	US-08-238- Sequence 5, Applicati	2.37e-02
C 8	24	1.8	108	2	US-08-878- Sequence 2, Applicati	2.46e-01
C 9	24	1.8	108	3	US-08-357- Sequence 2, Applicati	2.46e-01
C 10	24	1.8	965	3	US-08-388- Sequence 22, Applicati	2.46e-01
C 11	23	1.7	81	4	PCT-US95-1 Sequence 98, Applicati	7.69e-01
C 12	23	1.7	82	4	PCT-US95-1 Sequence 97, Applicati	7.69e-01
C 13	23	1.7	242	1	US-08-273- Sequence 1, Applicati	7.69e-01
C 14	23	1.7	776	5	5189151-1 Patent No. 5189151.	7.69e-01
C 15	23	1.7	965	3	US-08-388- Sequence 22, Applicati	7.69e-01
C 16	23	1.7	1800	5	5180581-1 Patent No. 5180581.	7.69e-01
C 17	23	1.7	2793	2	US-08-281- Sequence 5, Applicati	7.69e-01
C 18	23	1.7	2793	2	US-08-460- Sequence 7, Applicati	7.69e-01
C 19	22	1.6	74	4	PCT-US95-1 Sequence 100, Applicat	2.35e+00
C 20	22	1.6	74	4	PCT-US95-1 Sequence 100, Applicat	2.35e+00

C 21	22	1.6	75	4	PCT-US95-1 Sequence 99, Applicati	2.35e+00
C 22	22	1.6	81	4	PCT-US95-1 Sequence 98, Applicati	2.35e+00
C 23	22	1.6	156	1	PCT-US95-0 Sequence 2, Applicati	2.35e+00
C 24	22	1.6	156	1	US-08-021- Sequence 2, Applicati	2.35e+00
C 25	22	1.6	156	2	US-08-545- Sequence 2, Applicati	2.35e+00
C 26	22	1.6	156	1	US-08-084- Sequence 2, Applicati	2.35e+00
C 27	22	1.6	156	1	US-08-242- Sequence 2, Applicati	2.35e+00
C 28	22	1.6	156	1	PCT-US94-0 Sequence 2, Applicati	2.35e+00
C 29	22	1.6	162	1	US-08-242- Sequence 5, Applicati	2.35e+00
C 30	22	1.6	162	2	PCT-US95-0 Sequence 5, Applicati	2.35e+00
C 31	22	1.6	162	2	US-08-545- Sequence 5, Applicati	2.35e+00
C 32	22	1.6	540	2	US-08-117- Sequence 18, Applicati	2.35e+00
C 33	22	1.6	774	4	PCT-US91-0 Sequence 1, Applicati	2.35e+00
C 34	22	1.6	1011	4	PCT-US96-0 Sequence 119, Applicat	2.35e+00
C 35	22	1.6	1250	2	US-08-117- Sequence 59, Applicati	2.35e+00
C 36	22	1.6	1265	3	US-08-712- Sequence 5, Applicati	2.35e+00
C 37	22	1.6	1265	1	US-08-182- Sequence 5, Applicati	2.35e+00
C 38	22	1.6	1288	1	US-08-440- Sequence 9, Applicati	2.35e+00
C 39	21	1.6	4041	3	US-08-471- Sequence 22, Applicati	6.97e+00
C 40	22	1.6	4235	1	US-08-021- Sequence 3, Applicati	2.35e+00
C 41	22	1.6	4235	2	US-08-082- Sequence 3, Applicati	2.35e+00
C 42	22	1.6	4235	4	PCT-US94-0 Sequence 3, Applicati	2.35e+00
C 43	21	1.6	5852	3	US-07-867- Sequence 2, Applicati	6.97e+00
C 44	21	1.6	6049	3	US-08-471- Sequence 1, Applicati	6.97e+00
C 45	21	1.6	6049	3	US-08-463- Sequence 1, Applicati	6.97e+00

## ALIGNMENTS

RESULT 1  
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.  
AC xxxxxx  
DT Sequence 14, Application US/08232463  
DE Sequence 14, Application US/08232463  
CC Patent No. 5670367  
CC GENERAL INFORMATION:  
CC APPLICANT: DORNER, F.  
CC APPLICANT: SCHEIFLINGER, F.  
CC APPLICANT: FALKNER, F. G.  
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
CC NUMBER OF SEQUENCES: 52  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Foley & Lardner  
CC STREET: 1800 Diagonal Road, Suite 500  
CC CITY: Alexandria  
CC STATE: VA  
CC COUNTRY: USA  
CC ZIP: 22313-0299  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/232,463  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/935,313  
CC FILING DATE:  
CC APPLICATION NUMBER: EP 91 114 300.6  
CC FILING DATE: 26-AUG-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: BENT, Stephen A.  
CC REGISTRATION NUMBER: 29,768  
CC REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (703)836-9300  
CC TELEFAX: (703)863-4109  
CC TELEX: 899149  
CC INFORMATION FOR SEQ ID NO: 14:  
CC SEQUENCE CHARACTERISTICS:



DR WPI: 94-279739/34.  
 DR P-SDB: R65152.  
 PT Identifying proteins or peptide(s) which bind a ligand - by  
 PT screening a recombinant vector library expressing fusion proteins  
 PT comprising a binding domain and an effector domain  
 PS Disclosure: Page 35: 255pp; English.  
 CC 070466 is a generic DNA sequence used to generate random TSAR (Totally  
 CC Synthetic Affinity Reagents) peptides. This generic formula can also be  
 CC represented as follows: X(NNB)1(TGC)2(NNB)4Z(NNB)8(TGC)(NNB)  
 CC -97. X and Y are flanking restriction sites (X is not the same as Y)  
 CC that are not specified further. Other generic sequences are shown in  
 CC 070466-68. Other specific peptides generated by these generic sequences  
 CC are shown in R65151-54. TSARs are concatenated heterofunctional proteins  
 CC or peptides, comprising at least two functional regions - a binding  
 CC domain with affinity for a ligand and a second effector peptide portion  
 CC that is chemically or biologically active. They may further comprise a  
 CC linker peptide between the 2 domains. The oligonucleotides are also  
 CC designed so that the expressed peptide contains 2 or 4 cysteine residues  
 CC positioned in, or flanking, the unpredicted or variant residues. These  
 CC residues confer some degree of conformational rigidity to the peptides.  
 CC The TSARs or compns. comprising a TSAR binding domain can be used in  
 CC vivo to deliver a chemically or biologically active moiety, eg. metal  
 CC ion, radioisotope, peptide, toxin or enzyme, to the specific target or  
 CC on the cell. They can also replace the function of macromolecules, eg.  
 CC monoclonal or polyclonal antibodies and therefore circumvent the need for  
 CC complex methods of hybridoma formation or in vivo antibody production.  
 CC The TSARs are easily characterised and have designed activity allowing  
 CC direct and rapid detection in a screening process.  
 SO Sequence 114 BP: 0 A; 4 C; 4 G; 4 T;  
 Query Match 2.5%; Score 34; DB 12; Length 114;  
 Best Local Similarity 8.0%; Pred. No. 2,37e-04;  
 Matches 9; Conservative 30; Mismatches 73; Indels 0; Gaps 0;  
 Db 3 btgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62  
 Cp 1036 CTGCACACACTCTTATGATGATTTCTTCTGCTGTCATGTAGCACAATTT 977  
 Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114  
 Cp 976 CTCCTGATGCTGTCATATCATCTACATATACCTTGATTGCAATGTAGT 925

RESULT 15  
 ID V44650 standard; DNA: 91 BP.  
 AC V44650.1998 (first entry)  
 DT 06-OCT-1998  
 DE Mammalian DNA replication origin consensus sequence, uniorconsensus.  
 KW DNA replication origin; human; mammal; alphaconsensus; uniorconsensus;  
 KW anti-gene; DNA replication inhibitor; shuttle vector construct creation;  
 KW gene therapy; ss.  
 OS Mammalia.  
 PN WO9827200-A2.  
 PD 25-JUN-1998.  
 PF 12-DEC-1997; CA0972.  
 PR 21-MAY-1997; US-047322.  
 PR 16-DEC-1996; US-033374.  
 PA (UWMC-) UNIV MCGILL.  
 PI Cossens NH, Nielsen TO, Price GB, Zannis-Hadjopoulos M;  
 DR WPI: 98-362770/31.  
 PT Human or mammalian origin of replication consensus sequences - for  
 PT inhibiting DNA replication, for controlling initiation of  
 PT replication, maintaining circular plasmids and in assembly of human  
 PT artificial chromosomes  
 PS Claim 1; Page 42: 34pp; English.  
 CC This sequence represents a human or mammalian DNA replication origin  
 CC consensus sequences of the invention, designated uniorconsensus.  
 CC Administration of the consensus sequence or an anti-gene (comprising a  
 CC double stranded copy of the consensus) is used to inhibit DNA replication  
 CC in vivo or in vitro. The consensus sequences can also be inserted into an  
 CC expression vector, used subsequently for in vitro transfection of  
 CC mammalian cells, to control initiation of DNA replication. They can also  
 CC be used used to maintain circular plasmids that are capable of

CC semi-conservative replication in proliferating mammalian cells, or  
 CC inserted into mammalian or human artificial chromosome vectors for gene  
 CC therapy. Particularly, they are used to create shuttle vector constructs  
 CC for defining the essential mammalian elements required for maintenance of  
 CC chromosomal function. The consensus sequence can be combined with cloned  
 CC human telomeres and large centromeric blocks for assembly of human  
 CC artificial chromosomes and maintained as bacterial plasmids, circular or  
 CC linear, large or small yeast artificial chromosomes (YACs) or as episomal  
 CC elements.  
 SO Sequence 91 BP: 15 A; 1 C; 4 G; 7 T;  
 Query Match 2.4%; Score 33; DB 46; Length 91;  
 Best Local Similarity 12.2%; Pred. No. 6,91e-04;  
 Matches 11; Conservative 53; Mismatches 25; Indels 1; Gaps 1;  
 Db 2 wmtvaakravrwwkddavwagkrwwkvwvhrassacmddwkaaktwkgvltwarrwkg 61  
 Oy 186 TAAATACATAAAGAGATATGATGATATACAGTAAATGATGATCGAGTACCATGA 245  
 Db 62 -kmwttwawdatakwwkdkwkwrt 90  
 Oy 246 AGAATAGTACGTCACATCAATTGAATT 275

Search completed: Sat Nov 27 11:49:04 1999  
 Job time : 316 secs.



```

SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 2.7%; Score 36; DB 1; Length 204;
Best Local Similarity 12.5%; Pred. No. 2.7/e-05;
Matches 14; Conservative 53; Mismatches 45; Indels 0; Gaps 0

Db 73 aaycdchvgacgymrtthhyrimbdnyrdyrdsaaawccyrsvkydcynach 132
OY 505 AAACCTGATTTCGCAATTATGACAAATATTCATTCTCATCTGATTATTAAAGAT 564
Dd 133 ddhyvbybbvvnvhnncncbhnvchvbnhnhrnwayvrdhridv 184
OY 565 ATATATATGTTTGTGATGCATTCACGAAGATGCACATATATGTTAAAG 616

RESULT 13
ID ID Q51746 standard; cDNA; 91 BP.
AC Q51746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
KW Synthetic.
OS EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT ) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
PI WPI: 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PT Claim 3; Page 14; 23pp; English.
PS Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (051735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 2.5%; Score 34; DB 9; Length 91;
Best Local Similarity 11.7%; Pred. No. 2.37e-04;
Matches 7; Conservative 40; Mismatches 13; Indels 0; Gaps 0.

Db 12 svhyvvvvhvshvshvhhvhhvhsvvvhhvhhvhhvhhvhyvsvtcaagctcg 71
OY 855 GAACTAGATGTTTACAAATATTAAGAGCACTTGAAGACGTTTGCTGTTAACTCG 914

RESULT 14
ID ID 070466 standard; DNA; 114 BP.
AC 070466;
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR-9 peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key
FH misc-feature Location/Qualifiers
FT 55..60
FT /*tag-a
FT /note="this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"

```



	FW	Polypeptide; ORF: open reading frame; infection; bacterial;
KW	strepitococcal; bacteremia; diagnosis; prophylaxis; ds.	
OS	Streptococcus pneumoniae.	
FH	Key	
FT	Location/Qualifiers	
FT	CDS Complement (256..423)	
FT	/tag= a	
FT	/note= "polypeptide"	
FT	Complement (731..868)	
FT	/tag= b	
FT	/note= "polypeptide"	
PN	M09823631-A1.	
PD	04-JUN-1998.	
PE	24-NOV-1997: U21976.	
PR	27-NOV-1996: US-031879.	
PA	(SMIK ) SMITHKLINE BEECHAM CORP.	
PI	(SMIK ) SMITHKLINE BEECHAM PLC.	
PI	Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO,	
PI	Reid RH, Zarfos PN:	
DR	WPI: 98-32654/28.	
DR	p-PSDS: W62754, W63755.	
PT	Streptococcus pneumoniae polynucleotides - useful for developing	
PT	products for diagnosis, prevention and treatment of infections e.g.	
PT	pneumonia, bacteraemia, meningitis or endocarditis	
PS	Claim 1: Page 151-152; 181pp; English.	
CC	The sequence is that of a Streptococcal polypeptide coding region.	
CC	The polypeptide can potentially be used for the diagnosis and	
CC	prevention of bacterial infections, especially sp infection.	
CC	It may be used for the treatment of diseases such as otitis media,	
CC	conjunctivitis, pneumonia, bacteraemia, meningitis, sinusitis, pleural	
CC	empyema, endocarditis or infection of the cerebrospinal fluid.	
SQ	Sequence 1825 BP: 550 A: 385 C: 366 G: 524 T:	
	Query Match 11.6%; Score 157; DB 47; Length 1825;	
	Best Local Similarity 60.3%; Pred. No. 3,68e-72;	
	Matches 517; Conservative 0; Mismatches 336; Indels 4; Gaps 3;	
Dh	939 gagccataaatcttgcgcagataaacgacatctgttgtaagaatcggcgcaactcg 998	
Cp	1139 GAGCCAAAAATTACATAGAATAATACAGATCTTCTTTACATAACTTCGCAAAATCA 1080	
Dh	999 tcacaacagggaatggtcttcgttaaaagtatgcggtggaagaactctacaattccct 1058	
Cp	1079 TTTAATAATGCTTGCTGGTCTTAGAAGAGTGCGTGTTGAATA-CTGCACAACCTTTCTT 1021	
Dh	1059 gcttgygatattcttgcagcagccgcatccaagtcgcacaataattcttgttatgagc 1118	
Cp	1020 ATGTGGAATATTTCTTGCTGCTGCTGATCATATGTAGCACTAATTTCTCTTGATGCTGTC 961	
Dh	1119 aaagtcatacaatacaatcgtatcatacttgacaattttctcagtgaaacgaegttaacc 1178	
Cp	960 ATAATCACCTCAATATACCTTATTTGCCAATGTGATTTCAATTTGAAGAACGACCTTTMAACC 901	
Dh	1179 ggcaaatgttttcaagtcgacgcaccaagtccaatatcaatectctgcygtaaagaag 1238	
Cp	900 ACCAAACGCTTTCTAATGCTTTCTTAATATTTGTGAACATCTAGCTTCTCTAANTACTAAT 841	
Dh	1239 accaataacggcgtlccgcatcatgatatttgacaccaaagttgaaatgtaattg 1298	
Cp	840 CGCAATTACAGCTAATGATTTAAACAGTATGTGCACCAATATGTGGAAGACAGAGAGTG 781	
Dh	1239 ccccaagtttgcgaaggaatgaacgggtgaaggttgaaacagcttgttgaaacgaagaag 1358	
Cp	780 ATCATTAATAACCTCACCATTCACATACACATCAAAAAGCAGTACCTTATTCGGTAATTTGAAT 721	
Dh	1359 atcaactagctacaagaatcattgctcctgacctgaacaacataataataattgctatc 1418	
Cp	720 ATTGTGACATTAATGTCAT--CCGA-ATCTTTAAATCCATAGTAATTAATTTGGAATC 664	
Dh	1419 agaagtaattcttgaacaaatcagcatcttccacatagacaaaagaaccatcgtcaattg 1478	
Cp	663 TGCTTCAATTTTACGTATATGTTTCATCATCATACCCCAAGCAATAATACCTTTTAAATAT 604	
Dh	1479 tttyggcatgctgttaaaggcattgaaacatcctctgagacttgtaataatctgatatg 1538	

```

Cc      1599  atatcgcacagatccaagaacaaatatttggcattggccgaacacagactgtccatc 1658
Cc      483  ATATTTCACATCCCTCCAAAGGGAATATATCATCTTTCAGGCAATCCCATTCCTGTGCCATC 424
Db      1659  tccaatcaagaagcctgcatctgttaatgtagaagaacatgagacaacacatccgtcgt 1718
Cc      423  ACCANTTTAAATATGATGCTTTTATATCACCATTACATACATGATGATATAAACCTGTTGT 364
Cc      1719  tgaagtttttcacatgctgctctgcatctccatccatgcaacaaagcagcacaagaactac 1778
Cc      363  AGAAGTTTATACCATGTCGACACAGTTACAGCTACAGTATATGATTAATATTCGTGCC 304
Db      1779  tagaactcagcgtaac 1795
Cc      303  TAAAAAATCATTATTAAC 287

RESULT  10
ID      N81164 standard; DNA; 204 BP.
AC      N81164;
DE      08-NOV-1990 (first entry)
DT      Base substituted E.coli beta-galactosidase alpha-fragment.
KW      E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS      Escherichia coli.
FH      Key
FT      Location/Qualifiers
FT      misc.feature          .19..69
FT      primer_bind          187..204
FT      primer_bind          /*tag= a
                             /function= multiple cloning site
                             /*tag= b
PN      EP-285123-A.
PD      05-MAY-1988.
PR      30-MAR-1988; 105163.
PR      03-APR-1987; US-034819.
PA      (SUSO) SMOEN SOKERI OY.
PI      Leltovaara P, Knowles J, Koivuola A, Bamford J, Reinikainen T;
PI      WPI: 88-27927/40.
PT      Introducing random point mutations into nucleic acids -
PT      by prepn of single stranded template, annealing a primer, elongation,
PT      minicorporation, completion of molecules and screening.
PS      Disclosure; P: English.
CC      Random point mutations were introduced into the alpha fragment of
Cc      E.coli beta-galactosidase. The wild type sequence was obtained as a
Cc      single stranded, template and an oligonucleotide was hybridised to
Cc      it to generate a popn of DNA molecules which terminate at all
Cc      possible nucleotide positions within a specified region. The
Cc      variable 3' ends generated in this way are used as primers for
Cc      reverse transcriptase. Nucleotides are minicorporated by the
Cc      transcripase and the molecules are completed to forms that can be
Cc      amplified and then expressed in a suitable host-vector system.
Cc      The sequence covers all 176 diffit base substitutions, most of which
Cc      occurred singularly in any given mutant.
Cc      See also P80575.
SQ      Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match      3.0%; Score 40; DB 1; Length 204;
Best Local Similarity 6.5%; Pred. No. 3,19e-07;
Matches 6; Conservative 53; Mismatches 33; Indels 0; Gaps 0

Db      92  hhymrtdbnvrydrnsdaaawcyccrrsvkydcycnchddhyvbbvynvnhnnc 151
Cc      413  AATGAGAGCTTTTATATCACCATTACATAAACAAGTATATAAACCTGTTGTGAGAGTTTA 354
Db      152  ncccbnhvchvbnhbnhmvayrhaaridd 183
Cc      353  CCATGTGCACCATGTACAGCTACGAGAGTATA 322

```



CP 362 GAAGTTTACCATGTCACCACTACGCTACTGAGTATATGATCAATTAATCTGCTCT 303  
Db 1475 agaaactcattgtaacgcttgc-aggctgataccatttgctgcgaagaagcaattgcagc 1533  
CP 302 AAAAATATCATTTAATTAACATCAACATCTAATTTTCANTT-GATGTGACAGACATATTTCTTC 244  
Db 1534 gtgttattcgtcgaagaagcaattcccgagataattccataaccgcttagatttt 1593  
CP 243 ATGGCTA-CTCG-CGAAATG-CATTACCTTGTAATCACTACATCTCTTATGTTATT 187  
Db 1594 ttcatcaaaaggaagaagatgattccctgcgcctcaagaccgcttggtgaagtagta 1653  
CP 186 AGACCAATATGTAATATTTTATCCCTTATTTCTAGAGCACTCTGTAATATACGTA 127  
Db 1654 ctttcaaatctgataccctgaaccttgcccaatctgtgtaacaatcaagcccaagc 1713  
CP 126 GTTCATATATCCGATCCCTGTAACCTTCAATGCTTAATCATCATGATTTGTGTAATCA 67  
Db 1714 actcaccctgatacccttaattccgataaataatgata 1749  
CP 66 ACTCATGCCAGAACCTTAAATCCGACAAATGATA 31  
  
RESULT 8  
ID V27381 standard; DNA: 1267 BP.  
AC V27381;  
DT 02-OCT-1998 (first entry)  
DE Streptococcus pneumoniae sp0070 nucleotide.  
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
OS Streptococcus pneumoniae; otitis media; meningitis; ss.  
FH Key Location/Qualifiers  
FT CDS 2..1267  
FT /\*tag= a  
FT /product= "sp0070"  
FT /note= "no stop codon given"  
PD WO9818930-A2.  
PN 07-MAY-1998.  
PR 30-OCT-1997; U19422.  
PE 31-OCT-1996; US-029960.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Choi GH, Hiromocky J A, Johnson LS, Kunsch CA:  
DR WPI: 98-272224/24.  
P-PSDB: W551A0.  
PT Nucleic acid encoding antigenic peptide(s) from Streptococcus  
PT pneumoniae - or their epitope-containing fragments, useful in  
PT protective or therapeutic vaccines, and for diagnosis  
PS Claim 1; Page 72; 118pp; English.  
CC The present sequence encodes a protein from Streptococcus pneumoniae.  
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
CC can be useful in vaccines for inducing protective antibodies against  
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
CC are used to detect Streptococcus infection (by usual hybridisation or  
CC amplification methods), also for isolating Streptococcus genes or  
CC allelic variants. The protein can be used similarly to detect specific  
CC antibodies in standard immunoassays, especially for diagnosing or  
CC monitoring infections. Antibodies which bind the protein are used to  
CC detect corresponding antigens, to purify the protein and for passive  
CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
CC (especially 10-300) mu g/ml per dose.  
SQ Sequence 1267 BP; 365 A; 269 C; 261 G; 372 T;

Query Match 13.4%; Score 181; DB 46; Length 1267;  
Best Local Similarity 60.5%; Pred. No. 3,52e-86;  
Matches 637; Conservative 0; Mismatches 408; Indels 8; Gaps 7;

Db 11 gggacaaagctcggagctgagatgttgaagaagtactctctaccacaagcggtcttag 70  
Oy 94 GGACATGAGTTCAAGATCGATATTTGAGAACTACGATTTTACAGAAAGTTCCTTTAGA 153  
Db 71 cagcgaggaattacattcttcttcttgatgaanaaatctagaagcgatgatgaatt 130

Oy 154 AATAAGGGGATAAAAATATTTACATTTGCTGCTATTAACATAAAAAGAAATATGCTAGTT 213  
Db 131 atcgttggaatagccttcgcgcagatatcaacgfcgaatgtgctcctagcggaacaaat 190  
Oy 214 ATACAGGTAAATGCAAT-CG-CGAG-TAGCCATGAGAAATAGTAGTCACATTAAT-T 269  
Db 191 ggtacacatacaa-acgttacatgagttctgaagttacgttattagcttgacttgtag 249  
Oy 270 GAATTTAATGTTGTAATTTAATGATTTTATGAGACAGATTTATGATCAATATACCTC 329  
Db 250 catggagtagcaggaagacatggaanaaacttcaagcaggtatgtgtctatgctct 309  
Oy 330 AGTACCTGTAACTGCTGCACATGTGTAACACTTCTCAACAGAGTTTATATCAGATTTAT 389  
Db 310 gtctacattcaagttaccagcttccttgattggagatgagcaggtcgttggtcgcaa 369  
Oy 390 GAATGCTATAAAAAGACTTATTTTATTTATGATGGACACAGGTATGATGATGCTTGA 449  
Db 370 tgcacaataattgctcttgatactgcgaataatgagcgtcacttcaatgccttacacc 429  
Oy 450 AAGTATTTATTTCCGCTTTTGAGGCACTGTAATATAGAGTCACCTTTTAAGTTAATAAC 509  
Db 430 agaatactctattatcaccaaatgaccttgacacatccagattattcacaaagctcga 489  
Oy 510 TGATTAACCAATTTAGACAAATATGATTTGATCATCATCTGATTTTCAAAAGATTTAA 569  
Db 490 ggaatttttaagccttaacgactatgcaanaaatacccaaggctcttctgtcta 549  
Oy 570 TGATGTTTATGATGATTCCAAGAAATGGCACATATATGTTAAAAAGGATATATGCTTG 629  
Db 550 tggtaagatgctgaattgcgtaagattacgctcgtacacacatttataatgatt 609  
Oy 630 GGGTATATATGAACATCTACGTAATTAATGAGCAGATTTCCATTTTACTATGATT 689  
Db 610 tgaagctgaaggaatgaccttgtagctagtagtcttcttcgttcaataactggttcaac 669  
Oy 690 TAAAGATTC-GG--ATGACATTTATGCTCAAAATATTCAAATTCAGTAAAGTAGTCGC 746  
Db 670 ctccagcttatttcgctggaacaaacttggggcaattccacattccaacttgcgcg 729  
Oy 747 TTTGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 806  
Db 730 tcaacatctcgaatgagcaacgctgattgcttcttcaacaagaagattgatt 789  
Oy 807 CCATTAATTTTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 866  
Db 790 gaacttggcgtgagcaattgaaacatttgcggtgtgtaaacgcttccactgagaa 849  
Oy 867 TACAATATTTAAAGACATTAAGAAACGTTGGTGTTAAAGCGTTCATTAAGAAAC 926  
Db 850 aattgcaatgatacagtgattatcgatgacttggccacatccacaagaataattatgc 909  
Oy 927 TACAATTTCAAAATCAAGTATTTGATGATGATGATGATGATGATGATGATGATGATG 986  
Db 910 gaccttgatgcgctcgtcgaagaatccacaagaagaatcttagagccttcaacc 969  
Oy 987 TACAATTTACACACACACCAAGAAATATCCACATTAAGAAATGTTGTCAGATTTTCAAC 1046  
Db 970 gcaatccttcaagaacacattgcttggtagaagacttggccatgctttaaaccagc 1039  
Oy 1047 ACACACTTCTCTAGACACACAAAGCATTTTAAATGAATTTTCACAAAGTTTATGTAAGC 1106  
Db 1030 agatgctgttatactagcgaattatgctc 1062  
Oy 1107 AGATCGTGTATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1139  
  
RESULT 9  
ID V43027 standard; DNA: 1825 BP.  
AC V43027;  
DT 09-NOV-1998 (first entry)  
DE Streptococcus pneumoniae polypeptide coding region.



QY 91 TTAGGACATGAAGTTCAAGAGATCGGATATTGGAGAACTAGCTATTTCACGAAGTTGCCTT 150

Db 317 gaaaaatcaggtgtccctattttaccattatgacgataataatgataaagacatgatt 376

QY 151 AGAAATTAAGGGGATTAATAAATATTACCATTTGGTGTAAATTAACATAAAGAAAGATTATGTA 210

Db 377 gttattgcggaaatgcttttccagataccatgaagaatgcgccgcgcatcgaaatla 436

QY 211 GTTATACAAAGTAATGCAATTCGCGATAGCCAGTAAGAAATACTAGCTGCATCAATTG 270

Db 437 ggcgcgaagaatattcgttaccacagatttcattgctcgttttatacgaaacgttacaaagc 496

QY 271 AAATTAAGATGCTGTGAATTAATATATGATTTTATTTAGACAGATTAATTCATCAATTAATCA 330

Db 497 attgctgttaacagsggtcacatgaggaaacaagaagacagctggtttgctgacacatglatla 556

QY 331 GTAGCTGTAACTGTCGTCACATGTTAAACCTTCACAAACAGGTTTATATCAATGTTATG 390

Db 557 agtggatcaatccaaactagtatttaataagggagagcactgcccacgycgaacccanat 616

QY 391 AATGGATTAATAAAGACTTCATTATTTAATTTGGATGGCAGAGTATGGATTCCTCGAA 450

Db 617 gctgaatttccttgcaacttaagcgtgttgagatccgcgcattctctgtgcttaccac 676

QY 451 AGTGATTAATTTGCTTTTATGAGCATGTGATTAAGACGTACCTTTTAACTTATAAAGCT 510

Db 677 gattatgcattatgacgaaatalcgattttgatalccacagattactacaagaagacatgag 736

QY 511 GATTAAGCAATTAATGACAAATATTGATTTCGATCAATCCGTATTTATTCAAAGATATTAAAT 570

Db 737 gacgttttttcagcggttccaacaacatggtcatalcnaagtcnaaaaggaattttgcttat 796

QY 571 GATGTTTTTGGATGCAATTCACAGAAAGGCGCATATATGTTAAAAAAGCTATTATTGCTGG 630

Db 797 gttgtttgaagaatattcttgcagatagaatgcgaagtcgaatttatatttttggtgc 856

QY 631 GGTGATGATGAACATCTAGGTAAATTTGAAGCAGATGTTCCAAATTATTAATCTATGATATT 690

Db 857 agcgaagaagagkagatat 873

QY 691 AAGATTTGCGATGACAT 707

RESULT 7

ID V52194 standard; DNA; 11864 BP.

AC V52194;

DT 23-OCT-1998 (first entry)

DE Streptococcus pneumoniae genome fragment SEQ ID NO:61.

KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

KM computer readable medium; vaccine; pharmaceutical composition; ds.

OS Streptococcus pneumoniae.

PN M09818931-A2.

PD 07-MAY-1998.

PE 30-OCT-1997; U19588.

PR 31-OCT-1996; US-029960.

PA (HMDA-) HMDAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,

PI Kunsch CA, Rosen CA;

DR WPI: 98-272225/24.

PT Computer-readable medium with recorded Streptococcus pneumoniae

PT polynucleotide sequences - useful in diagnostic kits and assays, and

PT pharmaceutical compositions and vaccines for Streptococcus

PT pneumoniae

PS Claim 1: Page 524-530; 1409pp; English.

CC The present invention describes a computer readable medium which has

CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded

CC on it, or a representative fragment or a sequence at least 95% identical

CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1

CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus

CC pneumoniae. The present invention also describes an isolated nucleic acid

CC molecule encoding a homologue of any of the fragments of the S.pneumoniae

CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced

CC by a process comprising: (a) screening a genomic DNA library using as a

CC	probe a target sequence defined by any of the sequences in SEQ ID NO:1.
CC	To 391, identifying members of the library which contain sequences
CC	that hybridise to the target sequence and isolating the nucleic acid
CC	molecules from the members; or (b) isolating mRNA, DNA or cDNA produce
CC	from an organism; amplifying nucleic acid molecules whose nucleotide
CC	sequence is homologous to amplification primers derived from the
CC	fragment of the S. pneumoniae genome to prime the amplification and
CC	isolating the amplified sequences. The computer readable medium can be
CC	used in a computer-based system for identifying fragments of the
CC	S. pneumoniae genome of commercial importance, or expression modulat
CC	fragments of the S. pneumoniae genome. Products from the present
CC	invention can be used in diagnosis kits and assays, and pharmaceutical
CC	compositions and vaccines for S. pneumoniae.
CC	Sequence 11864 Bp; 3440 A; 2668 C; 2183 G; 3573 T;
So	
Query Match	15.2%; Score 206; DB 47; Length 11864;
Best Local Similarity	61.0%; Pred. No. 7, 05e-101;
Matches	661; Conservative 0; Mismatches 427; Indels 8; Gaps
Dg	635 gaggcataattgcgcgtagataaacacagcatcgtcttgltttaagaatcytgaagtgc 694
Cp	1139 GAGCGCAAAATTCACATAAAGATPACACGATCGCTTTCATTAACATTCTGCATATCA 1080
Dd	695 tccaacaaggcaatgfttccttctaagaatgatcggtttgaagaactcgtcacaaattcccttg 754
Cp	1079 TTGTAAAATGCTTGTTGTTCTAGAGAAAGTGCTGGTGAATACGCAACAACCTTCTTTA 1020
Dd	755 ctgggattcttcagcagcgccgcatccaagatcgcacaaattcttggttaagtcgtgcga 814
Cp	1019 TGTCGATATTTCTTCTGCTGCTGTTCAATTTGACACTAATTTCTTCTTGATGCTGTGCA 960
Dd	815 aagtcacgtgataacctgatatcatatgacaatttlctcagtgtaaagcagctttaaaccgcg 874
Cp	959 TAATCATCTCAATTAAGTAATGCAATTTGCAATTTGATGTTTCAATGAACGAGTTTAACCA 900
Dd	875 gcaaatgttttcaaagtctcagcaccgaattcaaatcaatcctgcgtgltgaagaaga 934
Cp	899 CCAACGTTTTCTATGCTTTCTTTAATTTGTATMACATCTAGCTTCTCTMAATTAACATAATC 840
Dd	935 ccaataacgcctgcgcgatcatcatgatattgtagcgcgaaggttggaaatgysgaattgc 994
Cp	839 GCATTTACAGTAAATGCAATTTAAACCTGTAATGCTCACATATTTGTGGACACGSAAGTA 780
Dd	995 ccacagitttgcacagcagaatagcagtgtagaaggttgaaccggttatbtgaacgaaga 1054
Cp	779 TCATAAATACACACATCCACATACATCAAAAGCATACCTTTATCCGTATTTTGATTA 720
Dd	1055 tcactagctcaaaatgcattgccttcgaagcttcaaaaaccaataaataatgttgatca 1114
Cp	719 TTTTAGACACATAAATGTCAT--CCGA-ATCTTTAAATCCATAGTAATTAATTTGAGAACATCT 663
Dd	1115 gagtaattttaagcaatttaagcatcttaacatlagacagaagaagcccttggttgtt 1174
Cp	662 GCTTCATATTTTACCTAATATTTTATATCATCCCAAGCATATATACCTTTTAAACATTA 603
Dd	1175 ttggcagatgctgttaaagggcatlaaaaaacatccctcgagactgtgaaataatcgtatgcg 1234
Cp	602 TGTCGCATTTCTTGATGATGCATCAAAACATCATTTAATCTTTGAAATTAACAGATGA 543
Dd	1235 tcaaatgtoaabtygttgatataagatattcctgggtgtaagcgaatgaagtcagctca 1294
Cp	542 TCGAAATCAATATTTGTCATTAATTTGGTATATCGGTTTAACTTTAAAAAGACGAGCTCA 483
Dd	1295 tattctgcagattcaagaacaaaattttggcatctggccgaaccaagacctgccatct 1354
Cp	482 TATTCACATGCTCTCAAAAAGGAAATATATACCTTTCAGGAATTCACATACCTTGTGCATCA 423
Dd	1355 ccaatacagaagctgcatctgtcaatgtgagaccaagaca tgaacaaca tactctgctt 1414
Cp	422 CCATTTAAATATGAAGCTTTTATATACCATTCATTAACATGTGATATATAAACCCTTTGTA 363
Dd	1415 gaagttttccatgtgctctgctactactccatgtcaacaagctcaagctcaaaagctact 1474



CP 812 GTATGTCACCATATTTGGAGACAGAGATGATCATATAAATCACCATCCACATACACA 753  
Db 601 tcaaaagcagt 611  
CP 752 TCAAAAGCAGT 742

RESULT 5  
ID X13717 standard; DNA; 677 BP.  
AC X13717;  
AT 19-MAR-1999 (first entry)  
DE Enterococcus faecalis genome contig SEQ ID NO:780.  
KM Enterococcus faecalis; contig: detection; Enterococcal infection;  
KW vaccine; attenuation; computer readable medium; ds.  
NS Enterococcus faecalis.  
OS MO9850555-A2.  
PD 12-NOV-1998.  
PF 04-MAY-1998; US-066009.  
PR 14-NOV-1997; US-066009.  
PR 06-MAY-1997; US-044031.  
PR 16-MAY-1997; US-046655.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Dillon PJ, Kunsch CA;  
DR WPI: 99-045171/04.  
PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
PT infection.  
PS Claim 1; Page 1983; 2084pp; English.  
CC A computer readable medium has been developed which has recorded on it  
CC 992 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC X12938 to X13919 represent these nucleotide sequences which are primary  
CC nucleotide sequences, also known as contigs. The computer-based system  
CC can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring  
CC progression of disease, and for identifying agents which can be used to  
CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
CC another related organism, in vivo or in vitro. In particular the  
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
CC can be used in vaccines to prevent or attenuate an Enterococcal  
CC infection.  
CC Sequence 677 BP; 204 A; 116 C; 142 G; 211 T;  
SQ

Query Match 16.8%; Score 227; DB 60; Length 677;  
Best Local Similarity 66.5%; Pred. No. 2.76e-113;  
Matches 446; Conservative 3; Mismatches 222; Indels 0; Gaps 0;

Db 7 tgaattcttgcatlgaagcggtgagatgacgcgtcaattcttggtctataccacaga 66  
OY 453 TGATATATTGCTTTGGAGCATGTGAATATAGACGTACCTTTTAAGTTAATACCGA 512

Db 67 ttatgcatatgacgaatattgattatcaccacgaattactacaagagcattgaga 126  
OY 513 TTACGCAATATGCAATATATTGATTCGATCCTGATTTTCAAGATATATATGA 572

Db 127 cgttttcagcgttcaaaacatggtcctcaagtcacaaagaaatattggtttatg 186  
OY 573 TGTGTTGATGTCATTCACAAATGGCACATATATGTTAAAGATATATGCTGGGG 632

Db 187 tgaatgaatattcttcgcagttagaaatcagaagtcgcagttatattatgagcgtac 246  
OY 633 TGATGATGACATCTAGTAAATATGAAGCATGTTCCATTTATATCTATGATTTAA 692

Db 247 gaagagggatgatataccaagccgaatattcaagaacacgaagcgtcatctttga 306  
OY 693 AGATTCCGATGACATTTATGCTCAAAATATCAATTAACGATTAAGGTACTGCTTTGA 752

Db 307 tgtttatcaagaagatattttagtgaatattggtcttaccagcatttggcctacaa 366  
OY 753 TGTGATGTCGATGATGATTTATGATCACTTCTGTCACAAATATGATGATGATAC 812

Db 367 tatmtgaagcgctaggttgattgctgctgctattttgrraaactgrrtagcmaaa 426  
OY 813 AGTTTAAATGCAATTAAGTCTAGTAAATGCGATTAGTATTATGAGAGCTAATGTTACAAA 872

Db 427 agtcgcagaagaatgcttaaggttgtaaacgctgcgtttttagcgagaanaagt 486  
OY 873 TATTAAAGAACCATTTACAAACGTTTGCTGTTAAACGCTTTTCATGAACTACCAAT 932

Db 487 cagtgacatgattatttgatgattatgacacacacacagctgaaatlaagaacgat 546  
OY 933 TGCAATTCAGATTATTTAGATGATTATGACACACATCCAAAGAAATTTAGTCTACAT 992

Db 547 tgaatgagctgcgcacaaatactgcacaaagaattatgctgcttccagccacatc 606  
OY 993 TGACACAGCAGCAAGAAATATCCATTAAGAAAGTTGTCAGATATTTCAACACACAC 1052

Db 607 attaacgacaaatctcctaagtgatgattgctgcaagcactgagatttgagatga 666  
OY 1053 TTCTCTAGAACACAAACATTTTAAATGAATTTGACAGAAAGTTATGTAAGCAGATCG 1112

Db 667 agtatcttat 677  
OY 1113 TGTATCTTAT 1123

RESULT 6  
ID X13228 standard; DNA; 4956 BP.  
AC X13228;  
AT 19-MAR-1999 (first entry)  
DE Enterococcus faecalis genome contig SEQ ID NO:291.  
KM Enterococcus faecalis; contig: detection; Enterococcal infection;  
KW vaccine; attenuation; computer readable medium; ds.  
NS Enterococcus faecalis.  
OS MO9850555-A2.  
PD 12-NOV-1998.  
PF 04-MAY-1998; US-066009.  
PR 14-NOV-1997; US-066009.  
PR 06-MAY-1997; US-044031.  
PR 16-MAY-1997; US-046655.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Dillon PJ, Kunsch CA;  
DR WPI: 99-045171/04.  
PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
PT infection.  
PS Claim 1; Page 1327-1330; 2084pp; English.  
CC A computer readable medium has been developed which has recorded on it  
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC X12938 to X13919 represent these nucleotide sequences which are primary  
CC nucleotide sequences, also known as contigs. The computer-based system  
CC can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring  
CC progression of disease, and for identifying agents which can be used to  
CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
CC another related organism, in vivo or in vitro. In particular the  
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
CC can be used in vaccines to prevent or attenuate an Enterococcal  
CC infection.  
CC Sequence 4956 BP; 1540 A; 927 C; 842 G; 1635 T;  
SQ

Query Match 16.6%; Score 224; DB 60; Length 4956;  
Best Local Similarity 66.2%; Pred. No. 1.65e-111;  
Matches 448; Conservative 1; Mismatches 228; Indels 0; Gaps 0;

Db 197 taccatttctggaatlaaggttctgcgatgacccattagcgttggcttaccacaa 256  
OY 31 TATCATTTTGTGGAATTAAGGTTCTGCGCATGATTCATTTAGCACAAATCATGATGAT 90

Db 257 aaagctacacgttcaagatcagacgttagaagatattcttaccacacggagatct 316



Query Match	48.98;	Score 660;	DB 60;	Length 660;
Best Local Similarity	100.08;	Pred. No. 0.00e+00;		
Matches 660;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

antimicrobial therapy; ss

Query Match	44.78;	Score 604;	DB 48;	Length 619;
Best Local Similarity	99.88;	Pred. No. 0.00e+00;		
Matches	610;	Conservative	0;	Mismatches 0; Indels 1; Gaps 1

Db	1	ctcatatataacatatataaagccatttctcagtcctaatcttctcaaatatgcat	60
Cp	1351	ctctattttttaaacaattttttaaagccatttttccatgcttaattttcctaaatgacattt	1292
Db	61	gtaatttttgaaatcacctgcaccataaataaacaacgagcttctcaattgctcta	120
Cp	1291	gtaatttttgaaatcacctgcaccataaataaacaacgagcttctcaattgctcta	1292
Db	121	atacataataagaatcttcataatgaagatgcaccccaatttcatcaataatct	180
Cp	1231	atacataataagaatcttcataatgaagatgcaccccaatttcatcaataatct	1173
Db	181	tgatcgcgttaatgcgcagaaattctcttaattgacgcaaaattccacataagaataca	240
Cp	1172	tgtatcgcgttaatgcgcagaaattctcttaattgacgcaaaatttccacataagaatca	1113
Db	241	cgatctgcttcatcaataaactttctgcaaatcatcttaaaaaatgcttgctctagagaaa	300
Cp	1112	cgatctgcttcatcaataaactttctgcaaatcatcttaaaaaatgcttgctctagagaaa	1053
Db	301	gtgctgcttgaataactgcgaacaactcttctatgctgatacttctctgctgctgtca	360
Cp	1052	gtgctgcttgaataactgcgaacaactcttctatgctgatacttctctgctgctgtca	993
Db	361	attgagacccaatttctcctgagtggtgcataatcatctacaataaacttgattgca	420
Cp	992	attgagacccaatttctcctgagtggtgcataatcatctacaataaacttgattgca	933
Db	421	attgtagcttcatggaacgacgctttaaaccacccaacgcttcttaagtcttctaata	480
Cp	932	attgtagcttcatggaacgacgctttaaaccacccaacgcttcttaagtcttctaata	873
Db	481	tttgtaacatctaagcttcttaataactaaatgcgaattacagcttaatgctttaaaact	540
Cp	872	tttgtaacatctaagcttcttaataactaaatgcgaattacagcttaatgctttaaaact	813
Db	541	gtatgctacatatgtgtagagacaggaagtgcataataaactcaacatccacatataca	600







CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or  
CC competitive polypeptide are useful for inhibiting the polypeptide e.g.  
CC bacterial (especially *S. aureus*) infections. They are also useful against  
CC *Helicobacter pylori* infections and related cancers, ulcers and gastritis.  
CC The antibacterial agents are useful to treat in-dwelling devices for  
CC infection prevention or generally as wound treatments to prevent adhesion  
CC of bacteria to matrix proteins. The Muc polypeptide is also useful for  
CC diagnosing or proposing a (susceptibility to) disease, for raising  
CC antibodies; to identify modulators or specific receptors; in rational  
CC drug design and as an immunogen for vaccines. The Muc gene sequences are  
CC useful in antisense/ribozyme therapeutics; to detect mutant Muc gene;  
CC for chromosomal mapping; to determine bacterial serotype; and for genetic  
CC immunisation.

50 Sequence 1351 BP, 471 A, 166 C, 260 G, 454 T;

Query Match	100.0%;	Score 1351;	DB 60;	Length 1351;
Best Local Similarity	100.0%;	Pred. No. 0.00e+00;		
Matches 1351; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

D	b	1	atggataaggagcttctataaagacacatcatcattcttcggaattaaaggtcttcg	60
O	y	1	atgagctaaaggattttatpatatcacactatcatcttctgctgaattaaaggtcttcg	60
D	b	61	atgagctcattgacacacaatcatgcatgatttaagacatgaagttccaaagtcg	60
O	y	61	atgagctcattgacacacaatcatgcatgatttaagacatgaagttccaaagtcg	60
D	b	121	gagaactacgtaattacagaagttgctccttagaaataaggagataaataattac	180
O	y	121	gagaactacgtaattacagaagttgctccttagaaataaggagataaataattac	180
D	b	181	gggtctataatacaataaagaagatagtgtgtctataaagtgtaatgcgagatgc	240
O	y	181	gggtctataatacaataaagaagatagtgtgtctataaagtgtaatgcgagatgc	240
D	b	241	catgaagaataagtagctgtgcacatcaatgaattagatgttgaagttaaatg	300
O	y	241	catgaagaataagtagctgtgcacatcaatgaattagatgttgaagttaaatg	300
D	b	301	ctgaagacagattatctgaatcaatactctcagtagctgtaactggtgcacatg	360
O	y	301	ctgaagacagattatctgaatcaatactctcagtagctgtaactggtgcacatg	360
D	b	361	ctctacaacaggtttatatcaacaatgctatgaagtgatgaataaagactcaatt	420
O	y	361	ctctacaacaggtttatatcaacaatgctatgaagtgatgaataaagactcaatt	420
D	b	421	ggataaggacaaggatgagagttcccgaaagtgataatttcgcttttgagcatg	480
O	y	421	ggataaggacaaggatgagagttcccgaaagtgataatttcgcttttgagcatg	480
D	b	481	tatagacgtcactctttaaagttataaaccctgatacgcataatgacaaatatg	540
O	y	481	tatagacgtcactctttaaagttataaaccctgatacgcataatgacaaatatg	540
D	b	541	gatcaaccctgattatctcaagaatataatgagtttttgatgcatccaagaatg	600
O	y	541	gatcaaccctgattatctcaagaatataatgagtttttgatgcatccaagaatg	600
D	b	601	catatgtttaaanaagttatatgtcttggtggatgataaactcagttaaattg	660
O	y	601	catatgtttaaanaagttatatgtcttggtggatgataaactcagttaaattg	660
D	b	661	gcagatgtccaatttatactatgatttaagaattcaagatccgatactttagct	720
O	y	661	gcagatgtccaatttatactatgatttaagaattcaagatccgatactttagct	720
D	b	721	atccaattacagataaagtgactgctcttgatgtgtaatgtaagtgatgttat	780
O	y	721	atccaattacagataaagtgactgctcttgatgtgtaatgtaagtgatgttat	780
D	b	781	caattcctgtctccaacataatggtgacatacagttttaaattcattagctg	840
O	y	781	caattcctgtctccaacataatggtgacatacagttttaaattcattagctg	840

QY	781	CACCTCTGCTCCACAAATNTGGTACCACTACAGTTTAAATGCATTACGTTAATTCG	840
Dp	841	attagttattttagaagaagctagatgcttcaacaatlatiaaagaagcattagaacgcttgg	900
QY	841	ATTAGTTATTTAGAAGAGCTAGATGTTACAAATATTTAAAGAAGCATTTAGAAACGTTTGT	900
Dp	901	gggtgttaaacgctcgcttccaatgaactacataatgcaaatcaagtatttctgtatgattat	960
QY	901	GGGTGTTAAAGCTGCTTCAATGAACCTCAAACTGCAAACTGCAAACTAATGTGATGATTAT	960
Dp	961	gcacacatccaagagaattagctctcaaatctgacagacagcagaagaatccacat	1020
QY	961	GCACACCATCCAAGAAATTTAGTCTACAAATTCACACAGCAGCAAGAAATATCCAT	1020
Dp	1021	aaagaagtcttgcaaglatltaacacacacacttctctctagaacacaagcatlltlaat	1080
QY	1021	AAAGAAGTTGTTGCAATATTTCAACACACACTTCTCTAGAACACAAGCATTTTAAAT	1080
Dp	1081	gaattgcagaaaagtttatgataagcagacatcgtgatacttctaigtgaatttggctca	1140
QY	1081	GAAATTCACAAAGTTTAGTAAACAGACAGTCTGTATTTCTTATGAAATTTTTGGCTCA	1140
Dp	1141	attagagaaaattcttgccgcatlaacgatacagaagaattaatctgataaatctgagatgca	1200
QY	1141	ATTAGAGAAATTTCTGCCCATTTACGATACAGATTAATTTGATTAATAATTTGGAGGTGCA	1200
Dp	1201	tcgtctctaatagaatccttlatatgataatgataagacaattgataatgctgtgttcta	1260
QY	1201	TCGTCATTAAAGATCTTTATTAATGATTAAGAACAAATTTGATTAATGCGTGTGTTTA	1260
Dp	1261	tttatgggtgcagtgatcatlcaaaaattcaaaaatgcatattagataaattagggcatg	1320
QY	1261	TTTATGGGTGCAGTGATTTTCAAAAATTTACAAAATGCATATTTTAGATAAATTAGCGCATG	1320
Dp	1321	aaaaatgcgttttaatatgtttataatagag	1351
QY	1321	AAAAATCGTTTTAAATATGTTTAAATATGAG	1351

```

RESULT 2
ID V74703; standard: DNA; 2424 BP.
AC V74703;
DE 16-MAR-1999 (first entry)
DT Staphylococcus aureus contig SEQ ID #392.
DE Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
Key Location/Qualifiers
misc_feature .1141..1200
                :/note="these bases represent a line of missing text in
                :/tag="a
                the sequence listing in the specification. They
                are included to maintain the nucleotide numbering
                given in the specification for this DNA sequence"
FN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
PI WPI: 97-374922/35.
DR Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 1287-1288; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so

```



\*\*\*\*\*  
Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
\*\*\*\*\*  
MURDER (TM)

\*\*\*\*\*  
Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
\*\*\*\*\*

MSrch\_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 11:43:48 1999; MasPar time 309.40 Seconds

Tabular output not generated. 935.614 Million cell updates/sec

Title: >US-09-103-287-1  
Description: (11351) from US09103287.seq  
Perfect Score: 1351  
N.A. Sequence: 1 ANGAGTAAGAGCTTTATAT.....TTAATATGTTATTAATAGAG 1351  
Comp: TACTCATCTCCAAATATA.....AATATACAAATATATCTC

Scoring table: TABLE default  
Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases. x 2

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database:

n-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39 40:part40 41:part41 42:part42 43:part43  
44:part44 45:part45 46:part46 47:part47 48:part48  
49:part49 50:part50 51:part51 52:part52 53:part53  
54:part54 55:part55 56:part56 57:part57 58:part58  
59:part59 60:part60

Statistics: Mean 9.075; Variance 6.414; scale 1.415

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1351	100.0	1351	60	UDP-N-acetylmuramate:	0.00e+00
2	1267	93.8	2424	58	Staphylococcus aureus	0.00e+00
3	660	48.9	660	60	Partial nucleotide se	0.00e+00
4	604	44.7	619	48	DNA encoding a Staphy	0.00e+00
5	227	16.8	677	60	Enterococcus faecalis	2.76e+13
6	224	15.2	4956	60	Enterococcus faecalis	1.65e+11
7	206	13.4	11864	47	Streptococcus pneumonia	7.05e+10
8	181	13.4	1267	46	Streptococcus pneumonia	3.52e+86
9	157	11.6	1825	47	Streptococcus pneumonia	3.68e+72

C	10	40	3.0	204	1	N81164	Base substituted E.co	3.19e-07
C	11	38	2.8	91	9	O51746	Oligonucleotide probe	2.98e-06
C	12	36	2.7	204	1	N81164	Base substituted E.co	2.71e-05
C	13	34	2.5	91	9	O51746	Oligonucleotide probe	2.37e-04
C	14	34	2.5	114	12	O70465	Generic DNA sequence	2.37e-04
C	15	33	2.4	91	46	V44650	Mammalian DNA replica	6.91e-04
C	16	32	2.4	114	12	O70465	Generic DNA sequence	2.00e-03
C	17	32	2.4	114	12	O70465	Generic DNA sequence	2.00e-03
C	18	32	2.4	114	12	O70465	Generic DNA sequence	2.00e-03
C	19	32	2.4	114	12	O70465	Generic DNA sequence	2.00e-03
C	20	30	2.2	91	46	V44650	Mammalian DNA replica	1.61e-02
C	21	29	2.1	501	3	N50028	Sequence encoding new	4.49e-02
C	22	29	2.0	114	12	O70470	Generic DNA sequence	3.36e-01
C	23	27	2.0	501	3	N50029	Sequence encoding new	3.36e-01
C	24	27	2.0	501	3	N50027	Sequence encoding new	3.36e-01
C	25	27	2.0	1200	2	O10187	Sequence encoding fun	3.36e-01
C	26	26	1.9	114	12	O70465	Generic DNA sequence	8.98e-01
C	27	25	1.9	114	12	O70470	Generic DNA sequence	2.36e+00
C	28	25	1.9	114	12	O70465	Generic DNA sequence	2.36e+00
C	29	26	1.9	114	12	O70472	Generic DNA sequence	8.98e-01
C	30	25	1.9	221	48	V38811	Homo sapiens CSP gen	2.36e+00
C	31	26	1.9	498	3	N50034	Sequence encoding new	8.98e-01
C	32	26	1.9	501	3	N50030	Sequence encoding new	8.98e-01
C	33	26	1.9	501	3	N50031	Sequence encoding new	8.98e-01
C	34	26	1.9	501	3	N50028	Sequence encoding new	8.98e-01
C	35	26	1.9	501	3	N50033	Sequence encoding new	8.98e-01
C	36	26	1.9	501	3	N50026	Sequence encoding new	8.98e-01
C	37	26	1.9	501	3	N50032	Sequence encoding new	8.98e-01
C	38	26	1.9	501	3	N50029	Sequence encoding new	8.98e-01
C	39	26	1.9	501	3	N50024	Sequence encoding new	8.98e-01
C	40	26	1.9	501	3	N50024	Sequence encoding new	8.98e-01
C	41	26	1.9	501	3	N50033	Sequence encoding new	8.98e-01
C	42	25	1.9	771	41	V17600	Staphylococcus aureus	2.36e+00
C	43	25	1.9	1167	44	T84187	DNA encoding an enoyl	2.36e+00
C	44	25	1.9	2479	51	V07906	Human cysteine-rich s	2.36e+00
C	45	25	1.9	2490	48	V38798	Homo sapiens cerebell	2.36e+00

#### ALIGNMENTS

RESULT 1  
ID V99650 standard; DNA: 1351 BP.  
AC V99650;  
17-MAR-1999 (first entry)  
DE UDP-N-acetylmuramate:L-alanine ligase (Murc polypeptide) encoding DNA.  
KW Murc gene; UDP-N-acetylmuramate:L-alanine ligase; Murc polypeptide;  
KW bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;  
KW Immunogen; drug; genetic immunisation; ds.  
OS Staphylococcus aureus.  
FH Location/Qualifiers  
FT 22..1335  
FT /\*tag= a  
FT /gene= "Murc"  
FT /product= "UDP-N-acetylmuramate:L-alanine ligase  
(Murc polypeptide)"  
EP-889123-A2.  
07-JAN-1999.  
PF 26-JUN-1998: 305064.  
PR 03-JUL-1997: US-052720.  
PA (SMRK ) SMITHKLINE BEECHAM CORP.  
PA (SMRK ) SMITHKLINE BEECHAM PLC.  
PI Burnham MKR, Wallis NG;  
DR WPI: 99-062655/06.  
P-PSDB: W87771.  
DR New isolated Murc polypeptide from Staphylococcus aureus and related  
PT nucleic acid - useful in diagnosis, treatment and prevention of  
PT bacterial infections  
PS Claim 2: Pages 3-4: 39pp: English.  
CC The present sequence represents a Murc gene encoding a Staphylococcus  
CC aureus UDP-N-acetylmuramate:L-alanine ligase (Murc polypeptide). Host  
CC cell containing an expression system comprising the Murc gene can be used  
CC for the recombinant production of the polypeptide. Agents or the Murc  
CC polypeptide are used to treat conditions requiring increased activity or



**THIS PAGE BLANK (USPTO)**



```

CDS      join(5276..5491,7810..7899)
          /gene="tat"
          /codon_start=5276
          join(5415..5492,7811..8104)
          /gene="rev"
          /codon_start=5415
          1..103
          /note="5' LTR"
CDS      257..1738
          /product="gag protein"
          /gene="gag"
          /codon_start=257
          1533..4542
          /product="pol protein"
          /gene="pol"
          /codon_start=1534
          /partial
          4487..5065
          /product="vif protein"
          /gene="vif"
          /codon_start=4487
          5005..5295
          /pseudo
          /product="vpr protein"
          /gene="vpr"
          /codon_start=5005
          5276..5491
          /number=1
          /gene="tat"
          5415..5492
          /number=1
          /gene="rev"
          5511..5758
          /pseudo
          /product="vpu protein"
          /gene="vpu"
          /codon_start=5511
          5676..8246
          /pseudo
          /product="envelope protein"
          /gene="env"
          /codon_start=5676
          7810..7899
          /number=2
          /gene="tat"
          7811..8104
          /number=2
          /gene="rev"
          8248..8865
          /product="nef protein"
          /gene="nef"
          /codon_start=8248
          8537..9000
          /note="3' LTR"
LTR      /note="3' LTR"
BASE COUNT      3265 a      1636 c      2208 g      2069 t
ORIGIN
Query Match      1.4% Score 19; DB 3; Length 9178;
Best Local Similarity 66.1%; Pred. NO. 3.83e+00;
Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
Db      2132 aaattcaaaattgggctgaataatcatatcactccagttatgctataaagaaa 2190
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp      1153 AATTTCCTAATGAGCCAAATTTCAATAGATACAGATCGCTTACATAAA 1095

```

Search completed: Sat Nov 27 11:49:34 1999  
 Job time : 13 secs.







```

allele      silent change"
              /product="reverse transcriptase"
              replace(262..264,"tgg")
              /note="reference sequence hxb2 has tgg, a tryptophan to a
              serine change"
              /product="reverse transcriptase"
              replace(346..348,"ttt")
              /note="reference sequence hxb2 has ttt, a phenotypically
              silent change"
              /product="reverse transcriptase"
              replace(361..363,"gat")
              /note="reference sequence hxb2 has gat an aspartic acid to
              a histidine change."
              /product="reverse transcriptase"
              replace(364..366,"gaa")
              /note="reference sequence hxb2 has gaa, a glutamic acid to
              a lysine change"
              /product="reverse transcriptase"
              replace(403..405,"ata")
              /note="reference sequence hxb2 has ata, an isoleucine to a
              threonine change"
              /product="reverse transcriptase"
              replace(407..409,"acc")
              /note="reference sequence hxb2 has acc, a phenotypically
              silent change"
              /product="reverse transcriptase"
              replace(484..486,"agt")
              /note="reference sequence hxb2 has agt, a phenotypically
              silent change"
              /product="reverse transcriptase"
              replace(493..495,"aca")
              /note="reference sequence hxb2 has aca, a threonine to a
              isoleucine change"
              /product="reverse transcriptase"
              replace(508..510,"cct")
              /note="reference sequence hxb2 has cct, a phenotypically
              silent change"
              /product="reverse transcriptase"
              replace(529..531,"gac")
              /note="reference sequence hxb2 has gac, an aspartic acid
              to a glutamic acid change"
              /product="reverse transcriptase"
              replace(541..549,"tac")
              /note="reference sequence hxb2 has tac, a phenotypically
              silent change"
              /product="reverse transcriptase"
              replace(598..600,"aca")
              /note="reference sequence hxb2 has aca, a threonine to an
              alanine change"
              /product="reverse transcriptase"
              replace(631..633,"agg")
              /note="reference sequence hxb2 has agg, an arginine to a
              lysine change"
              /product="reverse transcriptase"
              replace(679..681,"ttc")
              /note="reference sequence hxb2 has ttc, a phenotypically
              silent change"
              /product="reverse transcriptase"
              replace(736..738,"ctg")
              /note="reference sequence hxb2 has ctg, a phenotypically
              silent change"
              /product="reverse transcriptase"
              replace(784..786,"ggg")
              /note="reference sequence hxb2 has ggg, a phenotypically
              silent change"
              /product="reverse transcriptase"
              replace(799..792,"ttg")
              /note="reference sequence hxb2 has ttg, a phenotypically
              silent change"
              /product="reverse transcriptase"
              replace(811..813,"tac")
              /note="reference sequence hxb2 has tac, a phenotypically
              silent change"

allele      allele
              /product="reverse transcriptase"
              replace(856..858,"acc")
              /note="reference sequence hxb2 has acc, a threonine to a
              proline change"
              /product="reverse transcriptase"
              replace(877..879,"ata")
              /note="reference sequence hxb2 has ata, an isoleucine to a
              valine change"
              /product="reverse transcriptase"
              replace(889..891,"gaa")
              /note="reference sequence hxb2 has gaa, a glutamic acid to
              lysine change"
              /product="reverse transcriptase"
              replace(991..993,"aag")
              /note="reference sequence hxb2 has aag, a phenotypically
              silent change"
              /product="reverse transcriptase"
              replace(1030..1032,"gag")
              /note="reference sequence hxb2 has gag, a phenotypically
              silent change"
              /product="reverse transcriptase"
              replace(1156..1158,"act")
              /note="reference sequence hxb2 has act, a threonine to an
              isoleucine change"
              /product="reverse transcriptase"
              replace(1180..1182,"caa")
              /note="reference sequence hxb2 has caa, a phenotypically
              silent change"
              /product="reverse transcriptase"
              replace(1255..1257,"acc")
              /note="reference sequence hxb2 has acc, a phenotypically
              silent change"
              /product="reverse transcriptase"

allele      allele
              /product="reverse transcriptase"
              replace(1255..1257,"acc")
              /note="reference sequence hxb2 has acc, a phenotypically
              silent change"
              /product="reverse transcriptase"

BASE COUNT      531 a      217 c      263 g      288 t
ORIGIN
Query Match      1.4%; Score 19; DB 3; Length 1299;
Best Local Similarity 66.1%; Pred. No. 3,83e+00;
Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db 137 aaattcaaaattggcctggaatccataacatccagctatttctataaagaa 195
Cp 1153 AATTTCCTAATGAGCCAAAAATTTCACATTAAGAAATACAGATCTGTTACATAA 1095

RESULT 12
LOCUS      HIVNH51      1302 bp ss-DNA      VRL      22-SEP-1992
DEFINITION      Human Immunodeficiency virus type 1, clone NH51, partial pol cds.
ACCESSION      M00000
SEGMENT      1 of 3
SOURCE      Human immunodeficiency virus type 1 (HIV-1), clone NH51.
ORGANISM      Human immunodeficiency virus type 1
Virusidae; ss-RNA enveloped viruses; Positive strand RNA virus;
Retroviridae; Lentivirinae.
REFERENCE      1 (bases 1 to 1302)
AUTHORS      Weidt,G., Faistst,A.C., Sauermann,U., Lueke,W., Jentsch,K.D. and
Hunsmann,G.
TITLE      Characterization of a new German HIV-1 isolate: HIV-INHS
JOURNAL      Unpublished (1992)
STANDARD      full stat_entry
COMMENT      kindly provided prior to publication in computer-readable form by
Arne Faistst, Deutsches Primatenzentrum (DPZ), Abteilung fur
Virologie und Immunologie, Kellnerweg 4, W-3400 Goettingen,
Germany.
FEATURES
pept      from to/span      description
BASE COUNT      529 a      211 c      270 g      292 t
ORIGIN
Query Match      1.4%; Score 19; DB 3; Length 1302;
Best Local Similarity 66.1%; Pred. No. 3,83e+00;
Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

```



Query Match	1.5%;	Score 20;	DB 3;	Length 10276;
Best Local Similarity	67.2%;	Pred. No. 9.90e-01;		
Matches	39;	Conservative	0;	Mismatches 19; Indels

D6  
6768 tgaatctactcagaattcggcacaaatgtccacagaggcttttgatgcttgggatcat 6825  
||||| - | |||| ||||| - ||||| - ||||| ||  
QY 579 TGATGATTCCAGAAATGGCACATTAATGTTAAAAAAGTATTGTGCTGGGGTGAT 636

RESULT	9		
LOCUS	H1YM2S1	303 bp ss-RNA	VRL
DEFINITION	Human immunodeficiency virus type 1, clone 2, ADC sample M, partial gag cds, p17 region.		
ACCESSION	U01408		

ACCESSION	L21492
SEGMENT	1 of 4
SOURCE	Human immunodeficiency virus type 1 (HIV-1), clone 2, ADC sample M
ORGANISM	Human immunodeficiency virus type 1

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (pages 1 to 303)	Zhu, T., Mo, H., Wang, N., Nam, D. S., Cao, Y., Koup, R. A. and Ho, D. D.	Genotypic and phenotypic characterization of HIV-1 in patients with primary infection	Science 261, 1179-1181 (1993)

### FEATURES

Location/Qualifiers

```

source          <1...>303
                /organism="Human immunodeficiency virus type 1
BASE COUNT      125 a          61 c          62 g          55 t
ORIGIN          190 bp downstream from beginning of gag cds

```

	Query Match	Similarity	1.4%;	Score 19;	DB 3;	Length 303;
	Best Local	Similarity	77.1%;	Pred. No.	3.83+00;	
	Matches	27; Conservative		0;	Mismatches 8;	Indels 0; Gaps 0;
Db	82	aagatagatgttaaagacacccaagaagacttata	116			
Oy	856	AAGCTACATGTTACCAATATTAAAGACCATTA	890			

LOCUS	RESULT	10	VR	09-SEP-1993
DEFINITION	HIV10S1	309 bp ss-RNA		
DEFINITION	Human immunodeficiency virus type 1, clone 10, ADC sample F, partial gag cds, p17 region.			
DEFINITION	101356			

ACCESSION	L21375
SEGMENT	1 of 4
SOURCE	Human immunodeficiency virus type 1 (HIV-1), clone 10, ADC sample F
ORGANISM	Human immunodeficiency virus type 1
	Viridae: ss-RNA enveloped viruses; Positive strand RNA virus; Retroviridae: Lentivirinae.

REFERENCE	1 (pages 1 to 309)
AUTHORS	Zhu, T., Mo, H., Wang, N., Nam, D.S., Cao, Y., Koup, R.A. and Ho, D.D.
TITLE	Genotypic and phenotypic characterization of HIV-1 in patients with primary infection
JOURNAL	Science 261, 1179-1181 (1993)

**FEATURES**

Location/Qualifiers

patients. (A,M,L,F,C,V,R), in the study(1). All patients had acute, self-limited symptomatic illness with measurable viremia followed by seroconversion. Patient F was infected by patient M via sexual transmission. The authors report that the sequences in this study are found to be uniformly macrophage-tropic and noncytotoxic-inducing. The sequence for sample F clone 10, along with other patient F clones 1-9, 11, and 12, is shown in alignment following the printed text entry. These gap sequences cluster with HIV-1 B subtype sequences. See also L21224-L21591 and L24161, L24162.

FEATURES	Location/Qualifiers
CDS	<1..>309

BASE COUNT	128 a	64 c	62 g	55 t
ORIGIN	190 bp downstream from beginning of gag cds			
Query Match	1.4%	Score 19;	DB 3;	Length 309;
Best Local Similarity	77.1%;	Pred. No. 3,83e+00;		
Matches	27;	Conservative	0;	Mismatches 8;
			Indels	0;
			Gaps	0;

Db 82 aagatagatggttaaagaagacacccaagaagctttaga 116  
 ||| ||||| | | | ||||| |||||  
 QY 856 AAGCTAGATGTTACAATATTTAAAGAAGCATTAGA 890

RESULT	11			
LOCUS	HIVREF7..	1299 bp DNA	VRL	21-MAR-1994
DEFINITION	Human immunodeficiency virus type 1, isolate REF7, (E8RT) gen			
KEYWORDS	For reverse transcriptase			
ACCESSION	231355			
SOURCE	reverse transcriptase.			
ORGANISM	Human immunodeficiency virus type 1 (HIV-1), isolate REF7.			

REFERENCE	AUTHORS	TITLE	JOURNAL	STANDARD	COMMENT
1 (bases 1 to 1299)	Gurnandhe, A., Land, S., Birch, C., McAvay, C., Hooker, D., Trachetjan, G., Doherty, R. and Deacon, N.	Reverse transcriptase mutations and sequential HIV isolates from a patient with AIDS	unpublished (1994)	full automatic	Other sequences in this set are included in the 1994 Human

FEATURES	Location/Qualifiers
source	1..1299

allele

allele

allele



```

variant 5762 1 C in 4.41; t in B,P
variant 5798 1 t in 4.41; c in P
pept 5918 6223 vpr protein
6069 6358 tat protein, exon 2 (first expressed exon)
8597 8693 tat protein, exon 3 (AA at 8598)
variant 6208 1 t in 4.41; c in P
6213 6358 g in 4.41; a in B,P
6295 6358 rev protein, exon 2 (first expressed exon;
6597 8835 rev protein, exon 3 (AA at 8599)
variant 6296 1 t in 4.41; c in 1.5
6308 1 a in 4.41; g in 1.5
6343 1 c in 4.41; t in B,P
pept 6365 9034 env polyprotein
6407 1 t in 4.41; g in B,P
variant 6506 1 a in 4.41; g in 1.9
6556 1 a in 4.41; g in 1.9
6625 1 a in 4.41; g in B
6742 1 a in 4.41; c in 1.5
6749 1 c in 4.41; t in B,P
6789 1 c in 4.41; t in B,P
6911 3 a in 4.41; g in B,P
6955 1 a in 4.41; t in B,P
6977 1 a in 4.41; agcaatg in B,P
6978 1 g in 4.41; a in 1.9
6979 3 caa in 4.41; aag in B,P
variant 7048 1 a in 4.41; g in B,P
7163 1 t in 4.41; a in 1.9
7366 1 g in 4.41; a in B,P
7427 1 g in 4.41; a in 1.5
variant 7496 7 gagactg in 4.41; gatacta in B,P
7549 3 tct in 4.41; aca in B,P
7619 1 a in 4.41; g in 1.5, 1.9, B,P
7621 1 t in 4.41; g in 1.9
7625 1 c in 4.41; a in B,P
7738 1 g in 4.41; a in 1.5, P
7813 1 c in 4.41; t in B,P
7856 1 g in 4.41; c in B,P
variant 8024 1 a in 4.41; c in P
8112 1 a in 4.41; g in 1.5, 1.9, B,P
8150 1 a in 4.41; c in B,P
variant 8170 1 g in 4.41; a in B,P
8248 1 c in 4.41; g in P
8261 1 a in 4.41; a in P
8285 1 a in 4.41; t in 1.5, 1.9, B,P
8355 1 c in 4.41; g in B,P
variant 8423 1 a in 4.41; t in B,P
8477 1 g in 4.41; g in B,P
8597 1 g in 4.41; a in B,P
8679 1 g in 4.41; g in 1.5
variant 8758 1 a in 4.41; a in B,P
8775 1 a in 4.41; g in B,P
8784 1 c in 4.41; g in B,P
8849 1 c in 4.41; t in B,P
8855 1 c in 4.41; t in B,P
variant 8868 1 g in 4.41; a in B,P
pept 8898 9653 nef protein
8901 1 g in 4.41; a in B,P
8912 1 t in 4.41; a in B,P
8916 2 ta in 4.41; c in 1.5
8945 2 ta in 4.41; cg in B,P
variant 8972 1 g in 4.41; c in 1.5
8976 2 gg in 4.41; aa in B
9071 1 c in 4.41; t in B,P
9076 1 a in 4.41; g in B,P
9176 1 a in 4.41; g in B,P
9245 1 a in 4.41; t in 1.5
9267 2 ga in 4.41; at in B,P
9272 1 a in 4.41; g in B,P
9457 1 a in 4.41; g in P
9425 1 a in 4.41; g in B,P
9543 1 g in 4.41; a in B,P
9567 1 c in 4.41; a in B,P
9582 1 a in 4.41; g in B,P
9594 1 c in 4.41; a in B,P

```

```

variant 9642 1 a in 4.41; g in B,P
variant 9651 24 taagacagcagagcttcacaa in 4.41; tg in B,P
variant 9688 1 g in 4.41; t in 1.5, 1.9, B,P
variant 9705 1 c in 4.41; t in 1.5, 1.9, B,P
variant 9708 1 c in 4.41; t in B,P
variant 9746 1 g in 4.41; t in B,P
variant 9792 1 c in 4.41; t in P
variant 9870 1 a in 4.41; g in B,P
BASE COUNT 3414 a 1886 c 2469 g 2227 t
ORIGIN

Query Match 1.5%; Score 20; DB 3; Length 9996;
Best Local Similarity 71.7%; Pred. No. 9,30e-01;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Db 6559 agattgacattatcgtccagagcgttctgcttggaat 6604
||| ||||| ||||| ||| ||| ||||| |||
QY 591 AGAAATGGCACATTAATGTTAAAGATATTGCTTGCGTGAT 636

RESULT 8
LOCUS SIVSMH4 10276 bp ss-RNA VRL 01-MAR-1989
DEFINITION Simian (Sooty mangabey) immunodeficiency virus, H4 isolate,
complete genome.
ACCESSION X14307
SOURCE Simian (Sooty mangabey) immunodeficiency virus proviral DNA, clones
SMH3 and SMH4. Clone H41 is infectious.
REFERENCE 1 (bases 1 to 10276)
AUTHORS Hirsch,V.M., Olmstead,R.A., Murphy-Corb,M., Purcell,R.H. and
Johnson,P.R.
SIV from Sooty mangabeys: An African non-human primate lentivirus
closely related to HIV-2
Nature 339, 389-391 (1989)

TITLE
JOURNAL
STANDARD
REFERENCE
AUTHORS
JOURNAL
STANDARD
COMMENT
2 (bases 6391 to 6426; correction of [1])
Hirsch,V.M.
Unpublished, January 1990 (Hirsch, 301-496-2976)
Sequence kindly supplied in computer-readable form by V.M. Hirsch,
Georgetown University, NIH, Twindbrook II, 12441 Parklawn Drive,
Rockville, MD 20852 (02-FEB-1989).

The smh-4 sequence is presented in this entry with exception of
the extra "a" it has at position 4529 causing a frameshift in the
integrase of the pol gene. Clone smh-3 does not possess the
additional "a" and has a normal pol gene. Follow-up work revealed
the presence of a 36bp stretch -- coordinates 6391 to 6426 -- in
the smh-4 [ref 2]. Clone H41, without the extra 'a' in the integrase
cds, is infectious.

This sequence does not possess the premature stop codon in the env
cgs seen in other SIV sequences. It does possess an extended open
reading frame for nef.

The SM sequence differs from the macaque sequences by approximately
12% in the gag cds.

FEATURES
from to/span description
pept 1049 2572 gag polyprotein
pept 2338 5397 pol polyprotein (NH2 terminus uncertain)
pept 5327 5971 vif protein
pept 5799 6137 vpr protein
pept 6139 6444 vpx protein
pept 6290 6579 tat, exon 2 (first expressed exon)
pept 8806 8802 tat, exon 3 (AA at 8807)
pept 8516 6579 rev, exon 2 (first expressed exon)
pept 8806 9044 rev, exon 3 (AA at 8808)
pept 9243 9243 envelope polyprotein
pept 9077 10006 nef protein
pept 4528 4528 a in smh-3; aa in smh-4 (see COMMENT)
variant 6391 36 addition of 36 nt by [2]
BASE COUNT 3490 a 1903 c 2574 g 2309 t
ORIGIN
5' end of 5' LTR

```



```

YKHOEDKILKVGKFAKIKNTHTNGVSLAHVQKIGKEAIIIMGOVPRFLHVEKEI
WQOMWTDYQWVIMPEMDEVSTPLRVLFNFKVPEVIAEAFYFYDRSCNRSRGKA
GYVDRGREGKTLTLOTNOALEAFYALADSPCKANIYDSQVYNGVIGOTLES
ESRLVNOILEKMKKEKAIYAVNPAHKAIXENOYDHLVSQVIRYVLEKEIPQOE
HEKYSNVAKLVFKFLPRVAKQIVDTCDKHLGGEALHGVNAXLKGWDCDTHLE
GKIIIVAAHVSGFLEAEVIPPXETGRTALFLKLASRPVTHLTDNGANTSOEVK
MAVMAAGIEOTFXVYPNPOSQYVEAMNHHLTQIDRIEQNSIETIYLAHICINF
KRGGIGDMTPAERLVNMTTTEOEIOFQOSKSKFRNFRVYREGRDOLMGPGLELLM
KGEAVILKVGTEIKVYPRRKAIIKIDGCGKELDSGSLDEPTGAREVA"
4813..5457
/product="vlf protein"
/gene="vlf"
/codon_start=-1
/translation="MEEEKNWLIVPTMRIGRLERKMSHLIKLKNTKDLOKACYPH
HVVAMMTCSRVIFRDEALEVOGWNLTLPKXGMLSTVARTTWYSRRNWDATP
DYADLTLTGTYPCFSESEGYRAIAGEKILSCCKPKAKKNVPSLOIYALTVSHVX
SGEXPTMQRNRNRRGKRLATQXNRNKGSSSEFAGANFPGIAXKLGILA"
5285..5623
/product="vpX protein"
/gene="vpX"
/codon_start=-1
/translation="MSXPRERIPRNGSGEYGEAFDMHRTXEEINRAVNHLPREL
IQVNRKMEIWHDEGMSVSTKAKITLLOKALFMCKKCGCLDGGHGAGWRPG
PPPPPGILA"
5625..5930
/product="vpr protein"
/gene="vpr"
/codon_start=-1
/translation="MTERPEDEAPOREMDEWVYVLEEKELANKFDPRLIALG
NHYHGDGTLAGELIRILGALFIHRGGRSHRSIGSGXNPLSTIPXKGVL"
5776..6066
/number=1
/gene="tat"
/partial
6002..6066
/number=1
/gene="rev"
/partial
6072..8732
/product="env protein"
/gene="env"
/codon_start=-1
/translation="MGCLGNOLLALSLASAGIYCVQYVYGVAPRNATVPLFC
ATRNDRDTGTCLOCPDNDYSELAIVTEAPFAMPTVEAIEVMNLFERSIPCV
KLPICITMRCKNKKSEXXRMLGXGAPATPTOSTPPSIKAVVNDSPCIRSNCT
GLEOBMVSCCKNMTGKDRKREYNETWYSDDLYCEQNSNENEDSCYMHCTSY
IOESCKRHWDAIRERYCAPPGYALRCNDTISGAPNCTIVVSSCTRMETDST
WGFNGTRENRTYIYMHRSNRTIISLKNYXNLMRCRPNKTVLPVTIMSGVFFH
SOPINERKQAMCWEGGEWEKAIREVXXLVHPKRGKXKXOIKLTPAGGDEVT
FMWTCRGEFLCYCKMNFVLMYEXIKNGSRMTSONOKERYONVPCHXOIIIMHK
VGNVXLPREDLDCNSTVSTLAEIDMNGKNITMSAVALRELDGYLVE
ITDGPAPSVKRYTTTGASRNKRGVYGFGLATAGASAGASALITSASRLLA
GIVOOOOLLDVYKROELRLTWGTMLOTRVATEYKLDQAPXNSWGFARQVC
HTVPEPNDTLTPXNNXMXQMEKQVNFLENTIXYLEAQIQEEXNMYELQKLNXX
DXFGNMXDLTIXKIYGVGLVLYIGLVIYVOMIARLRQGRVFFSSPAXVX
QIPITGDELPTKEGEGDGGXRGGRSWPQIEXIHFIHRIOLRLTMFLFSSCDML
LXCOILOVLOLSRTIORAREVIREITYQYGRVYQEAQAQWKKXKRETLASAW
RLMELTGRVGGIILAIRIROGLETLL"
8296..8391
/number=2
/partial
8566..9351
/product="nef protein"
/gene="nef"
/codon_start=-1
/translation="MGVTSKQKQAGNLAERLLQARGETVGRIMWEGIEGYSSQXD
ASGKGLSLSCPEKXCEGQFNNTPRNATGAKIYQOONMDDVNDNDIVGCVSP
RPIRVIMTYKLAIDMSHFIKEGLEGYYSIRRHIIIDLYEKEGILIPDMQNTAG
PDIRYPMFGMLMKLXPVNVSDQEDETHLMPHAPOTSDMPDGEVLAMKFDXLA
YXKAFVHEBEEFGSSGSKREYORRLRLARLLIKMADKKKS"
BASE COUNT      3252 a 1785 c 2350 g 2143 t 108 others
Origin      Start of 5' R repeat.

```

```

Query Match      1.5%; Score 20; DB 3; Length 9638;
Best Local Similarity 71.7%; Pred. No. 9,90e-01;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Db 6266 agaattggcaatlaatgctacagagctttgagcttgagataat 6311
Oy 591 AGAATGCGACATATGTTAAAGGATATGTTGCTTGCGGTGAT 636

RESULT 7
LOCUS SIVSMPDBJ 9996 bp ss-RNA VRL 03-JAN-1990
DEFINITION Simian (sooty mangabey) immunodeficiency pbj virus, isolates
bc13 and bc11; complete genome. Infectious clones.
ACCESSION M31325
SOURCE SIVSYM (sooty mangabey) isolates bc13 and bc11. The prototype
infectious 4.41 was constituted from PCR fragments 4 and 41 of
bc13; for other clones, see COMMENT.
REFERENCE Dewhurst, S., Embretson, J.E., Anderson, D.C., Mullins, J.I. and
Fultz, P.N.
TITLE Sequence analysis and acute pathogenicity of molecularly cloned
SIV-smpbj14
JOURNAL Nature 345, 636-640 (1990)
STANDARD
COMMENT Kindly submitted in computer readable form prior to publication by
James Mullins, Stanford University, Palo Alto, CA.

The PBJ virus is the most acutely pathogenic primate lentivirus
reported to date. The prototype shown below is a chimera derived
from two PCR fragments; other chimeric clones of bc13 are 1.9
(infectious) and 1.5 (noninfectious), possibly due to a T -> C
change in the rev start codon. Of the bc11 isolate, the B and P
clones are both nonpathogenic.

PBJ-4.41 has the second NFKappa-B element found in HIV-1's but not
in HIV-2's nor in the SIV's related to the HIV-2's, including the
other known sooty mangabey isolate, H4, nor in the nonpathogenic B
and P clones (see annotation at position 9651 below and discussion
in [1], Fig. 2).

FEATURES
        from to/span      description
variant 36 1 a in 4.41; g in 1.5
variant 239 1 c in 4.41; t in 1.5
variant 525 1 c in 4.41; g in 1.5
variant 724 1 c in 4.41; a in 1.5
variant 771 1 c in 4.41; ca in 1.5
pept 828 2351 gag polyprotein
variant 873 1 c in 4.41; c in 1.5
variant 915 1 c in 4.41; a in 1.5
variant 1115 1 c in 4.41; t in 1.5
variant 1205 1 a in 4.41; g in 1.5
variant 1622 1 a in 4.41; t in 1.5
variant 1635 1 c in 4.41; g in 1.5
variant 1742 1 c in 4.41; ctc in 1.5
variant 1751 3 tct in 4.41; t in 1.5
variant 1751 3 a in 4.41; t in 1.5
pept 2108 5176 pol polyprotein
variant 3232 1 g in 4.41; a in 1.5
variant 3264 1 c in 4.41; g in 1.5
variant 3514 1 c in 4.41; g in 1.5
variant 3601 1 a in 4.41; t in 1.5
variant 3841 1 a in 4.41; c in 1.5
variant 3901 1 a in 4.41; g in 1.5
variant 4062 1 a in 4.41; g in 1.5
variant 4561 1 t in 4.41; c in 1.5
variant 4812 3 aag in 4.41; ttc in 1.5
pept 5106 5750 vlf protein
variant 5198 1 t in 4.41; c in B,P
variant 5297 1 a in 4.41; g in B,P
variant 5525 1 t in 4.41; c in P
pept 5578 5916 vpX protein
variant 5653 1 g in 4.41; c in B,P
variant 5700 1 g in 4.41; a in B,P
variant 5706 1 a in 4.41; g in B,P
variant 5746 1 c in 4.41; t in 1.5

```



CDS		325..>582 /note="vpu protein" /translation="MIDLLAKVDYRLAANAFLIATIFALVWTVIAIEFRKLIRQK IDRLIERIEREDSGNESDGDITLSTVDRGNLRLLDAVDV 503...>3031 /note="env polyprotein" /translation="MRVMGIOMNCCOOMWIGGILGFEMLLMGNGMGLAWTVYYGVVPV WQEAHPLELCASDAQAYEREHNVNVAETTLACVPLDPSPOLMENLVENFNMKNKVAV DOMHDDISLMDEBGLKPCVKYLPIPLCTTACCSNKVATGTAGSYTNTSEDMRNCFN ITTELRLDKRKRYALEFKLDIVPLNGENEYRIINONTSSITIQACKRVFEPIPIHY CAPAVYAIIICRDITKENGPGCKNVSTVCGTHIKPVSTOLLNSTAEEDIIINS EMLYNNAKIILVOLNQSVGINCTRPNNRSGRIRIPGOTFAATGDIIDIOAHON ISROKWEKLOOVKGKLEHFKNFKTIREAFNSGGSDIETHSHFGCEFFCYCSTALF INDSNAPINNSSDTANITLCCRIOKITIMMOEGVAMAPLOGIIFAKNSNTGL LTRDGSTNSTDETFRFGGDDMDNRSELYKKVYIARLGLAPTRAKRVVERERK AALAVALGFELGAAGSTWGAASITTLVVARQLLSGIVQOOSMLKAIEVQHMLDLTT VWGIGLOLQTRVLAIERYLKDOOLLGITWGSGLICTTAVPNMSSMSNKSQEITWDNM TMWMDREINTNYETIIRLLEYSOEQOEENEDLLALDNOMLSWFEDITKMLWTIK IFIMIVGLGILGRIFAVALSVINVRQSGSPISFOLLINPPGPDRPGRIDEEGGGO DKGRSTRPLVSFGIALAMDRLRSFLSFYSRLDILLIYARVEILGCRMEPLTKYG SLYQVWGLELKRSAVSLDITTAIAVAEGIDRIITEIQRIMRAPCNIPRRIRGFEAA LQ"
CDS		3036..>3340 /note="nef protein" /translation="MGNKWSKGMPAVREIRRTKPAREVNAAAEGVGAASODLDKY GALTISNPANNADCAWMLETQEEBEVEFPVPVPLPRMTITKGAFDSLFTLEKGG L"
BASE COUNT	1191 a    568 c    791 g    790 t	
ORIGIN	226 bp upstream from the beginning of rev cds start	
Query Match	1.5%; Score 20; DB 3; Length 3340;	
Best Local Similarity	70.8%; Pred. No. 9,90e-01;	
Matches	34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;	
Dc	1642 tacatcagcccttgtaataatgaattacaataatgaccacattaataa 1689	
Cp	1305 TAAATATGCATTGTTGTAATTTTGAAATACACCTGCACCATAAATA 1258	
RESULT	5 HSPUENV 4549 bp ss-RNA VRL 15-MAR-1988	
LOCUS	Human spumaretrovirus (HFV or foamy virus) genomic RNA with env gene.	
DEFINITION	Human spumaretrovirus (HFV or foamy virus)	
ACCESSION	X05591 Y00070 Human spumaretrovirus HFV.	
SOURCE	1 (bases 1 to 4549); enum. 1 to 4549)	
REFERENCE	Fluegel,R.M., Rethwilm,A., Maurer,B. and Darai,G.	
AUTHORS	Nucleotide sequence analysis of the env gene and its flanking regions of the human spumaretrovirus reveals two novel genes	
TITLE	EMBO J. 6, 2077-2084 (1987)	
JOURNAL		
STANDARD COMMENT	Recombinant clones that represent the 3' part of the genome of the human spumaretrovirus (foamy virus) were established from viral DNA (clone C55: see X05592) and cDNA (clone B52). The region common to both viral inserts comprises 817 bp (from the BamHI at bp 4548 in X05591 to the HindIII site at bp 816 in X05592 correspondy to bp 5365 in a fused-oned viral sequence. The env precursor polypeptide contains 14 pot. N-linked glycosylation sites. *source: cell type=HSRV-infected HEL fibroblasts; clone=B52;	
FEATURES	Data kindly reviewed (02-SEP-1987) by Fluegel R.M. Location/Qualifiers CDS <1..1032 /note="3' domain (integrase) of reverse transcriptase; po gene product (343 AA)" /codon_start=1 1171..4128 /note="env glycoprotein precursor (AA 1-985)" /codon_start=1171 1525 a 812 c 885 g 1327 t	
BASE COUNT	1525 a 812 c 885 g 1327 t	
ORIGIN	1.5%; Score 20; DB 3; Length 4549;	
Query Match		

Db	4281	ttatcatgataataagaataggagacaacatcctcaacatataaactgcaagat	4338
Matches	39	Conservative	0; Mismatches 19; Indels 0; Gaps 0;
Db	678	ttactatgatgatttaagaatcgcatgacatttatgctctcaaaatttcaaatatgacgat	735
RESULT	6		
LOCUS	SIVSM9	9638 bp ds-DNA	VRL 09-MAR-1992
DEFINITION	Simian immunodeficiency virus, sooty mangabey isolate SM9, complete genomic sequence.		
ACCESSION	M80194		
SOURCE	Simian immunodeficiency virus (sooty mangabey) isolate SM9 (P. Fultz, Univ. of Alabama) DNA.		
ORGANISM	Simian immunodeficiency virus		
REFERENCE	1 (bases 1 to 9638)		
AUTHORS	Cournaud, V., Laure, F., Fultz, P.N., Montagnier, L., Brechot, C. and Sonigo, P.		
TITLE	Genetic differences accounting for evolution and pathogenicity of simian immunodeficiency virus from a sooty mangabey after cross-species transmission to a pig-tailed macaque		
JOURNAL	J. Virol. 66, 414-419 (1992)		
STANDARD	full staff-entry		
COMMENT	The SM9 isolate when passaged through a rhesus macaque gave rise to the highly infectious SIVSM9P (pb14) virus (see <SIVSM9P>, M31325 and M80193). The authors estimate that a maximal set of 57 nonsynonymous nucleotide substitutions and two insertions (one in the LTR, the other in the env cds) account for the acutely lethal phenotypic change.		
FEATURES			
CDS	Location/Qualifiers join(5776..6065,8295..8391) /product="tat protein"		
	/gene="tat"		
	/codon_start=1		
	/translation="MEMPILKQESLSSRHSISSEVDADTPESASLEBETLSOLX RPNQACINKCTCKCKCHQCFILKKGIGICEORRRTPKNTANFASDKLSYK ARNQCPKREKEVEYEVATDGLNR"		
	join(6002..6065,8295..8533) /product="rev protein"		
	/gene="rev"		
	/codon_start=1		
	/translation="MGSNEELRRRLRILHFLHOTNIFYXXGPGTANORRRRRRKROR WQQLALADRIYSPDPADPTPLDLAIQOGLAIEKLPPNPASAPELKDAESP"		
	1..310 /note="R+U5"		
	535..2058 /product="gag protein"		
	/gene="gag"		
	/codon_start=1		
	/translation="MGVRSVLSGKKADELEKILRPGCKKKYMLKHIVAAANELDRFR GLSEILKEGCKRLSVLAPVPTGSENLKSLYVRYVRLMCHAEKXKHEEAKO IVOHVITVETADKMPATSRPRAPRGSGNPVVOVGNYTHLPSPTLNAVYKL VEEKFAEYVPGQALSEGTPLDIOIMNCVGHQAAQIILEINLEPADADLQH POPOPPIAGQLREPRGDIAGTSTXQEOIQMWRQONPAPGVITYRWLQIGLOKCV RMVNPINILPVKQGPKEFSQYDREKFSRAEDTDAKXNMOTLIQNLNANDCKL VLKGIINPTLEMLTACOGYGGGPKXRLMAELKDPLNOGPLFAVVOOKGXKIL XCKNCKEHSARQFRAPRROGCKWCGKAGHVAKCPEKQAGTGLGPKKKPNPFM AQMFGITPTAPDPDPAVDLKNMKYGRRORENRPRYKEVTEIDLHLNSLGEEO"		
	1815..4883 /product="reverse transcriptase"		
	/gene="pol"		
	/codon_start=1		
	/translation="MPKKTGFPRAMPKGEAPQFPHGPASGADNTCSKRGSSCGST ELHEBGGQKBEQRETLQGGNGGFAPOFSLMRPVYVAYIEBQPLEVLLDGDADS IVAGIELGPNYTPRIWVGIGIFINTKIKYKVKILVKIKGIMTGDTEINFGRL LTMAGMLNPIAKVEPIKYLKRGKGPRLQMPLSKERIILAREICEMEKDQGLE XAPPTNNTPTFAIKKXNDKRMMLIDPRLKNKVPQDEEVOLGIPHPGLAKRRI TVLDVGAFTSPIDEFEROYTAFTLPSVNAABGKRYIKVYLPQGGKSPALFOHXY RNLPEPRKANPDTYLIQYDDILIASDRDLDEHRYVLDKELNSIGSTPEKQK KDPFOMWGEIEMPTKWLQKILDEPQETIVNNIQLVGLVWMAAIVYGITKIKLCS RLRGKTKTLEEVQWTEMAAEAYEENKIPISOEGEGYOEGRPLEAPYKSDONWS		



```

site      8782      8784      in-frame stop in env cds [5]
site      8785      8787      in-frame stop in env cds [1],[2],[3],[4],[6]
refnumbr  506       506       numbered 1 in [1]
signal    10102     10107     mRNA polyadenylation signal
variant   517       519       cgg in [6]; gc in [1]
variant   617       618       tg in [6]; t in [1]
variant   668       668       g in [6]; gc in [1]
variant   694       694       g in [6]; c in [1]
variant   779       779       a in [6]; g in [1]
variant   806       806       a in [6]; agt in [1]
variant   825       826       ac in [1],[6]; gag in [4]
variant   843       844       ga in [4],[6]; agaa in [1]
variant   858       859       ct in [4],[6]; cct in [1]
variant   870       870       a in [1],[6]; g in [4]
variant   895       897       gag in [1],[6]; gc in [4]
variant   919       919       g in [4],[6]; a in [1]
variant   966       968       gct in [1],[6]; gt in [4]
variant   1009      1011      atc in [4],[6]; ac in [1]
variant   1014      1014      g in [1],[6]; a in [4]
variant   1056      1056      t in [4],[6]; g in [1]
variant   1083      1083      g in [1],[6]; a in [4]
variant   1190      1190      g in [4],[6]; c in [1]
variant   1365      1365      g in [1],[6]; a in [4]
variant   1396      1396      c in [1],[6]; t in [4]
variant   1435      1435      g in [1],[6]; a in [4]
variant   1619      1619      g in [4],[6]; a in [1]
variant   1848      1848      g in [6]; a in [1]
variant   1882      1882      c in [6]; t in [1]
variant   1905      1905      g in [6]; a in [1]
variant   2156      2158      cgc in [6]; gcg in [1]
variant   2170      2189      t in [6]; c in [1]
variant   2190      2190      ga in [6]; ac in [1]
variant   2277      2277      g in [6]; a in [1]
variant   2336      2336      g in [6]; a in [1]
variant   2717      2717      a in [6]; g in [1]
variant   2747      2747      g in [6]; a in [1]
variant   2858      2858      a in [6]; g in [1]
variant   3053      3053      g in [6]; c in [1]
variant   3083      3083      g in [6]; ga in [1]
variant   3104      3104      ag in [6]; a in [1]
variant   3702      3702      a in [6]; c in [1]
variant   3648      3648      g in [6]; t in [1]
variant   3839      3839      t in [6]; a in [1]
variant   3878      3878      g in [6]; a in [1]
variant   3971      3973      tgg in [6]; gta in [1]
variant   4131      4131      g in [6]; t in [1]
variant   4218      4218      a in [6]; t in [1]
variant   4231      4237      agaacg in [6]; gaca in [1]
variant   4261      4261      a in [6]; g in [1]
variant   4266      4266      t in [6]; a in [1]
variant   4271      4271      c in [6]; ta in [1]
variant   4283      4284      ac in [6]; a in [1]
variant   4302      4302      c in [6]; g in [1]
variant   4318      4320      ttc in [6]; att in [1]
variant   4324      4325      at in [6]; agt in [1]
variant   4350      4350      atg in [6]; tt in [1]
variant   4405      4407      aaa in [6]; a in [1]
variant   4413      4415      aaa in [6]; gg in [1]
variant   4439      4442      caca in [6]; ctga in [1]
variant   4470      4474      accac in [6]; gtcca in [1]
variant   4490      4490      a in [6]; t in [1]
variant   4786      4790      aagac in [6]; a in [1]
variant   4804      4804      t in [6]; tc in [1]
variant   4832      4832      a in [6]; tat in [1]
variant   4850      4851      ga in [6]; cag in [1]
variant   4870      4870      c in [6]; a in [1]
variant   4892      4892      g in [6]; a in [1]
variant   4911      4911      a in [6]; ag in [1]
variant   4921      4922      tg in [6]; t in [1]
variant   5560      5560      t in [6]; c in [1]
variant   5599      5599      aa in [6]; gag in [1]
variant   5632      5632      a in [6]; g in [1]
variant   5639      5639      a in [6]; tc in [1]

```

```

variant 5660 5662 ttg in [6]; t in [1]
refnumbr 5731 5731 numbered 1 in [3]
variant 5737 5737 a in [1],[4],[6]; g in [3]
variant 5817 5818 tg in [1],[4],[6]; aa in [3]
variant 5877 5877 g in [1],[4],[6]; a in [3]
variant 5895 5895 g in [1],[4],[6]; g in [1],[3]
variant 5968 5968 g in [1],[4],[6]; a in [3]
variant 6070 6070 g in [4],[6]; a in [3]
variant 6078 6078 a in [2],[3],[6]; g in [4]
variant 6078 6078 numbered 1 in [2]
refnumbr 6143 6143 c in [4],[6]; t in [2],[3]
variant 6143 6143

Note: remainder of annotations omitted.

Query Match 1.6% Score 21: DB 3: Length 10277;
Best Local Similarity 75.6%; Pred. No. 2,31e-01;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 6774 agaattggcccttaattgtacagaagcttgatgctggg 6814
Qy 591 AGAATGGCACAATATCTTAAAGATGATTCCTTGGG 631

RESULT 4
LOCUS HIVDJ373A 3340 bp ds-DNA VRL 31-DEC-1993
DEFINITION Human immunodeficiency virus type 1, DJ373 from Djibouti,
proviral DNA encoding env, tat, vif, rev, and nef genes.
L23065
KEYWORDS env gene; envelope glycoprotein; nef gene; rev gene; tat gene;
vif gene.
SOURCE Human immunodeficiency virus type 1 (HIV-1), DJ373 from Djibouti,
PCR amplification of proviral DNA from cocultivated PBMC's
HIV-1
ORGANISM Human immunodeficiency virus type 1
Virus: ss-RNA enveloped viruses; Positive strand RNA virus;
Retroviridae; Lentivirinae.
REFERENCE 1 (bases 1 to 3340)
AUTHORS Louwagie,J.J., Janssens,W., Mascola,J.J., Fischer,C.L., Hegerich,P.A.,
van der Groen,G., McCutchan,F.F.E., Eddy,G. and Burke,D.
TITLE Genetic diversity of the HIV-1 envelope glycoprotein
JOURNAL J. Virol. 68(1994)1000-1008
COMMENT full automatic
STANDARD
These sequences were kindly provided prior to publication by Dr.
Joost Louwagie of the Henry M. Jackson Foundation Research Laboratory
Rockville, Maryland. Twenty-one full length gp160 coding sequences
from eight African countries (Djibouti, Gabon, Kenya, Senegal,
Somalia, Uganda, Zaire, and Zambia) were sequenced and analyzed with
thirty-two previously published full-length gp160 env sequences.
Sequences DJ258, DJ259, K124, SE365, SM145, UG266, UG274,
VI191, and VI525 have corresponding gag sequences published by
Louwagie et al. in AIDS 7, 769-780 (1993) and in the 1993 compen-
dium. This sequence clusters with C subtype env sequences. See also
accession numbers U22939-U22957 and L23064.
NCBI gi: 437346

FEATURES
source
1..3340
Location/Qualifiers
/organism="Human immunodeficiency virus type 1"
/isolate="Djibouti"
/cell_type="T-lymphocyte"
/proviral
/sequenced_mol="DNA"
/tissue_type="blood"
join(87..>302,2619..>2707)
/note="tat protein, exon 2 (first expressed exon)"
/translation="MEVPDNPDLPMWHPGQPTACKCKCKCKCHCLVFOYKGL
GISYGRKRKRORRSAPSSSEDHONLISKQ"
/note="tat protein, exon 3 (aa at 2620)"
/translation="PLSRGQDPTGPEESKRVKSTKADPLD"
join(226..>302,2619..>2838)
/note="rev protein, exon 2 (first expressed exon)"
/translation="MAGRSQSDSDALQAVRIKILHQ"
/note="rev protein, exon 3 (aa at 2621)"
/translation="SPYPEKGTROARORNRARRARORORIHISIERISTLCGRPA
EVPFOLPIERLIDCSSESGTSGTEGVN"

```







```
CDS /codon_start=4826
5298..5636
/note="vpx protein"
/codon_start=5298
5637..5942
/note="vpr protein"
/codon_start=5637
5788..6083
/note="tat protein, exon 2 (first expressed exon)"
6014..6083
/note="rev protein, exon 2 (first expressed exon)"
6084..8300
/note="tat cds intron 2"
6084..8300
/note="rev cds intron 2"
6084..8300
/note="tat, rev, nef subgenomic mRNA intron 2"
6090..8738
/pseudo
/note="env polypeptide (in-frame stop codon at 8298)"
/codon_start=6090
8298..8300
/note="in-frame stop in env cds"
8301..8397
/note="tat protein, exon 3 (AA at 8302)"
8301..8554
/note="rev protein, exon 3 (AA at 8303)"
8572..9360
/note="nef protein"
/codon_start=8572
8957..9646
/note="3' LTR"
9471..9646
/repeat_region
/note="R repeat 3' copy"
9624..9629
/misc_signal
/note="mRNA polyadenylation signal"
BASE COUNT 3290 a 1818 c 2386 g 2152 t
ORIGIN Cap site of genomic RNA.

Query Match 1 6%: Score 21; DB 3; Length 9646;
Best Local Similarity 75.6%; Pred. No. 2,31e-01;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 6284 agaatggccctaatgtacagaagcttgatcgagg 6324
||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 591 AGAATGGCACAATATGTTAAAAAGTATTTCTCTGGG 631

RESULT 2
LOCUS SIYVMA11 10274 bp ss-RNA VRL 18-OCT-1991
DEFINITION Simian immunodeficiency virus from the same rhesus macaque that
SOURCE provided the virus for the clone SIYVMA231; independent molecular
ACCESSION M76764
SOURCE M76764
REFERENCE 1 (bases 1 to 10274)
AUTHORS Luciw, P.A., Shaw, K.E., Unger, R.E., Planellas, V., Stout, M.W.,
Pitt-Lowe, E., Leung, N.J., Banapur, B. and Marthas, M.L.
TITLE Genetic and biologic comparisons of pathogenic and non-pathogenic
JOURNAL molecular clones of simian immunodeficiency virus (SIVmac)
COMMENT AIDS Res. Hum. Retroviruses (1991) In press

Kindly provided to GenBank in computer readable form prior to
publication by R.E. Shaw. See also <SIYVMA239> (accession number
M33262); SIYVMA11 and SIYVMA239 show greater than 98% homology at
the nucleotide level. Analysis in [1] largely focuses on sequence
differences in SIYVMA11 and SIYVMA239 which may account for
differences in biologic properties.
Location/Qualifiers
Join(6301..6596,8799..8898)
/note="for splice patterns, see Unger, R.E. et al."

FEATURES
CDS
```

```
CDS /product="tat"
/codon_start=1
/translation="METHPLEQENSLSSGSHSSCSSEAASTPELANLGEITLSOLY
RPLEACTNTCTCKKCTHCQCFKLGKGLGCTCYOSRRKRTPKAKANSTSSANNRPI
SNRRHCQPEKAKETVEKAVATAPGLGR"
Join(6527..6596,8799..9055)
/note="for splice patterns, see Unger, R.E. et al."
/product="rev"
/codon_start=1
/translation="MSHEREEELRRRLRIHLLHOTIDPPTPGPGTANORRRRRR
RRRQDLALADRIYSFPDPTDPLDALIQQLNALAESIPDPPTNTPEALCDPTKD
RSRPOD"
1..817
/note="5' LTR"
487..492
/note="R = 518"
1052..2584
/product="gag"
/codon_start=1
/translation="MGARNSVLGKKADELEKIRLPNGKKKMYLKHVYAAANELRF
GLAESILENREGGOKITSVLAPVPGSENLKSLTYCYINCINIAEKKKHTPEAKO
ITQRLHYEVETGRLAEMPKTSRPTAPSSGREGNPVOOIGGNTYVHLPLSPRTLAWL
IEBKKEGAEVVPGFQALSEGCTPYDINQMLNCVGDQAAQIIRDIINEAADMWDLQ
POPAPQOGLREPSGSDIAGTSSVDQIQMAARQONIPVGNIRYRIQGLQKCVR
MYNPNTLIDYKOGPEFOSYVDRFYKSLAEQTDAAVKMMTQTLIONANPDKLY
LKIGVNPTELEMLTACGGVGGPGRKARLMAELKALAPVPIPPAAOORGRPKIK
CMNCKRGHAROCRAPRRRCCKKCKGKMDVMAKCDPROGFIGLGPMGKPPNPFMA
QVHGGLPTIAPPEDPADVLKNTYQJAKQOREKRSREKPYKEMTEDLILNLSLGG
DO"
2227..5409
/product="pol"
/codon_start=1
/translation="MLEMERGLTCKRAMSPKTKMILEMKNPCYQGMPTRGFFR
PMSGKRAPFPFGSSASGADANCSPRGSCGSAKEIHAGQAEKTEKQEBALOG
DDEGFAAPQFSLMRRPVYTAHIEGQPEVLDTGADSLVTGLELPHYKPLVGGIG
GFLNTEKYNVEIEVLGKRIJGTIMGDTPIINFGNRLTALQMSINLPIAKVEPKV
ALPBGKVPKIKOMPLSKERIKVALREICEEMDQGLEAPPTNPNYKTFIAKKDK
NMKRMILDFRELNRVODTEVOLGIDPHGKLAKRRIRIVLDIGDAVFSIPLDEPFO
YTAFTLPSVNNAGPRKRYIKYKLPQCKSGSPALFOYTMRRVLEPRKANDPTLYOM
DDLILASDRNLNEDRVLDLQELNLSIGSTEEKFOKPPFOMMGIELMPTKIKQ
KIELPQREITVNDIQVLGVNLWMAQIIPGITKHLCRILRKMTLITELVQTEMAE
AEYENKRIISQOEGCYOEGKPLEVATYIKNDONMSYRIHQEDMILKVGKPAKIKN
THNGVYLLAHVIOIGKEAIVIMGPKEHLPEKDYOMOMTIDWQVYIEMNFI
STPPIVATVNLKPDPIEGEETYYTDSCKKSKGKAGYITRGGDWVYLEOTNO
QALBERALMALTSGRANITVDSQYVMGIIITCPESESLVNLQIKKIKSEIY
AWPFAHKGIGNEVDHLVSQIRQVLELEKIEPADQEHDKYSNKEVLEKGLPEYI
VARQIYDTCRCHQKGEALHGOVNSDGTQNMOCITLHGKRIIVAVASGFIYAEVI
POETGROTALFLKLAGRMPITHLHTDNGANFASOEKVMYAMAGIEHTGCVYNPDS
OGVYEAAMHHLKNOIDRIEIOANSVEIYIMAVHSMNFKRGGTGDMTPERLNMIT
TEOIEIOOSKNSKFKNFRYRYREGRODLKGGELLMKGEAVILKVGTDIVVPRR
KATIIDYGGGKEMDSSNHEDTGEAREVA"
5339..5983
/product="vif"
/codon_start=1
/translation="MEEEKRMAVPTWRIPERLERMSHLIKYLKTKDLQKCYVPH
FKVGMAMWMTSVIPLPQBSHLVNGVYMLTPEKGLSYVAARVYVSNFMDTTP
NYADILHSTYFPCTTAGEVYRAIRGQILSCCRFPRAKHGYVPSLOYLAKVYSVR
SGDENPTWKQWRDNRGLRMARONSGDKRGKGPPTKANPGLAIGILA"
5811..6149
/product="vpx"
/codon_start=1
/translation="MSPRRRIRPGNGEETIGEAPEMLNRTYEINREAVNHLPEL
IFQWQSRWMEYHMDQGMSPSYKRYRLCTIQRLALFWCKKGRCLGEGGAGGWRPG
PPEPPPGLA"
6150..6455
/pseudo
/note="premature stop @ 6185"
/product="vpr"
/codon_start=1
6603..9239
/pseudo
/note="Accession Number M74144(1 to 1572 = 6603 to 8174;
SU)"
```







**THIS PAGE BLANK (USPTO)**



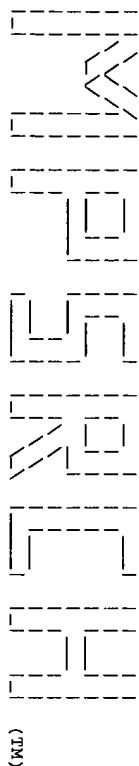




Query Match	1.98; Score 127; DB 1; Length 211;
-------------	------------------------------------

	Query Match	1.9%	Score 127;	DB 1;	Length 497;
	Best Local Similarity	51.1%;	Pred. No.	2.72e+00;	
	Matches	23; Conservative	11;	Mismatches 11;	Indels 0; Gaps 0;
Dd	k i e v k d t k e a l e k i				
Dc	aaarathtgargtmaargayacmaragcnungaraarathg 313				
OY	854 AGAAGCTGAGATTACCAATTTTAAAGCATTGAACGGTTTG 898				
dt	K L D V T N I K E A L E T F				





(TM)

Release 3.1a John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mparch\_ntp n.a. - n.a. Smith-Waterman search, using a protein database  
which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Nov 24 01:07:08 1999; MasPar time 2.75 Seconds  
Tabular output not generated. 1143.653 Million cell updates/sec

Title: >US-09-103-287-1  
Description: (1-1351) from US09103287.seq  
Perfect Score: 6755  
N.A. Sequence: 1 ATGAGTAAAGAGCTTTATAT.....TTAATAGTTTATATAGAG 1351  
Comp: TACTCATCTCTCAAAATATA.....AATTATACAAATATATATCTC

Scoring table: TABLE backtranslated  
Gap 30

Mmatch STD : Dbase 0; Query 0

Searched: 1479 segs, 1162203 bases x 2  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: HIV-AA8  
1:ALL

Statistics: Mean 77.190; Variance 217.128; scale 0.356

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	137	2.0	91	1 GAGSM2S1	HIVM2S1, GAG	5.31e-01
2	137	2.0	103	1 GAGSF10S1	HIVF10S1, GAG	5.31e-01
3	135	2.0	486	1 GAGSVI69	HIVVI69, GAG POLYPROT	7.44e-01
4	137	2.0	491	1 GAGSPH136	HIVPH136, GAG POLYPRO	5.31e-01
5	137	2.0	500	1 GAGSCAM1	HIVCAM1, GAG POLYPROT	5.31e-01
6	127	1.9	211	1 GAGSBZ200	HIVBZ200, GAG POLYPRO	2.72e+00
7	126	1.9	489	1 GAGSVI557	HIVVI557, GAG POLYPRO	3.18e+00
8	131	1.9	493	1 GAGSBZ167	HIVBZ167, GAG POLYPRO	1.44e+00
9	127	1.9	497	1 GAGSFBI321	HIVFBI321, GAG	2.72e+00
10	127	1.9	500	1 GAGSVU2	HIVVU2, GAG POLYPROTE	2.72e+00
11	127	1.9	500	1 GAGSBZ33C	HIVBZ33C, GAG POLYPRO	2.72e+00
12	127	1.9	500	1 GAGSVU10	HIVVU10, GAG POLYPROT	2.72e+00
13	121	1.8	486	1 GAGSBV231	HIVBV231, GAG POLYPRO	6.77e+00
14	121	1.8	486	1 GAGSVI141	HIVVI141, GAG	6.77e+00
15	121	1.8	492	1 GAGSVI205	HIVVI205, GAG POLYPRO	5.84e+00
16	122	1.8	493	1 GAGSVI455	HIVV455, GAG POLYPROT	5.84e+00
17	122	1.8	502	1 GAGSEF2	HIVSEF2, GAG POLYPROTE	5.84e+00
18	122	1.8	647	1 ENVSD747	HIVD747, ENV POLYPROT	5.84e+00
19	120	1.8	738	1 ENV82GHI	HIV2GHI, ENV POLYPROT	7.84e+00

20	120	1.8	846	1 ENVELOPE52	HIV2ISY, ENVELOPE POL	7.84e+00
21	120	1.8	846	1 ENVELOPE52	HIV2ISYR, ENVELOPE PO	7.84e+00
22	120	1.8	851	1 ENV82D194	HIV2D194, ENV	7.84e+00
23	120	1.8	859	1 ENV82ST	HIV2ST, ENV POLYPROTE	7.84e+00
24	120	1.8	859	1 ENV82CAM2	HIV2CAM2, ENV POLYPRO	7.84e+00
25	118	1.7	75	1 GAGSL6S1	HIVL6S1, GAG	1.05e+01
26	118	1.7	101	1 GAGSC4S1	HIVC4S1, GAG	1.05e+01
27	115	1.7	147	1 REV8AGM677	STVAGM677, REV PROTEI	1.59e+01
28	117	1.7	264	1 GAGSVI354	HIVVI354, GAG POLYPRO	1.21e+01
29	117	1.7	304	1 GAGSVI159	HIVVI159, GAG POLYPROT	1.21e+01
30	113	1.7	385	1 GAGSVI132	HIVVI132, GAG POLYPROT	2.09e+01
31	117	1.7	486	1 GAGSVI415	HIVVI415, GAG POLYPRO	1.21e+01
32	116	1.7	487	1 GAGSR89	HIVR89, GAG POLYPROTE	1.39e+01
33	115	1.7	490	1 GAGSVI203	HIVVI203, GAG POLYPRO	1.59e+01
34	116	1.7	491	1 GAGSE365	HIVSE365, GAG POLYPRO	1.39e+01
35	116	1.7	492	1 GAGSK124	HIVK124, GAG POLYPROT	1.39e+01
36	117	1.7	495	1 GAGSSM145	HIVSM145, GAG POLYPRO	1.21e+01
37	117	1.7	497	1 GAGSPH153	HIVPH153, GAG POLYPRO	1.21e+01
38	117	1.7	499	1 GAGSOY1	HIVOY1, GAG POLYPROTE	1.21e+01
39	116	1.7	499	1 GAGSHAN	HIVHAN, GAG POLYPROTE	1.39e+01
40	118	1.7	506	1 GAGSMN	HIVMN, GAG POLYPROTEI	1.05e+01
41	116	1.7	643	1 ENVSD757	HIVD757, ENV POLYPROT	1.39e+01
42	116	1.7	650	1 ENVSD760	HIVD760, ENV POLYPROT	1.39e+01
43	115	1.7	848	1 ENVELOPE_U	HIVU08455, ENVELOPE	1.59e+01
44	116	1.7	851	1 ENVELOPE_U	HIVUJF, ENVELOPE POLY	1.39e+01
45	115	1.7	852	1 ENVELOPE_U	HIVU08453, ENVELOPE	1.59e+01

## ALIGNMENTS

RESULT 1  
ID GAGSM2S1 PRELIMINARY; PRT; 91 AA  
AC L21492  
DT 09-SEP-1993  
DE HIVM2S1, GAG  
FT PEPT  
CC -1-5':N 3':N  
CC TRANSLATED USING PHASE 1  
SQ SEQUENCE 91 AA; 10138 MW;

Query Match 2.0%; Score 137; DB 1; Length 91;  
Best Local Similarity 51.1%; Pred. No. 5.31e-01;  
Matches 23; Conservative 12; Mismatches 10; Indels 0; Gaps 0;  
Db k l d v k d t k e a l e k i  
Dt 80 araratthgagynaargayacnaargargcnytnararathg 124  
QY 854 AGAGCTAGATGTTACCAATATTAAAGACCATTAAGCGTTTG 898  
Qt K L D V T N I K E A L E T F

RESULT 2  
ID GAGSF10S1 PRELIMINARY; PRT; 103 AA  
AC L21375  
DT 09-SEP-1993  
DE HIVF10S1, GAG  
FT PEPT  
CC -1-5':N 3':N  
CC TRANSLATED USING PHASE 1  
SQ SEQUENCE 103 AA; 11552 MW;

Query Match 2.0%; Score 137; DB 1; Length 103;  
Best Local Similarity 51.1%; Pred. No. 5.31e-01;  
Matches 23; Conservative 12; Mismatches 10; Indels 0; Gaps 0;  
Db k l d v k d t k e a l e k i  
Dt 80 araratthgagynaargayacnaargargcnytnararathg 124  
QY 854 AGAGCTAGATGTTACCAATATTAAAGACCATTAAGCGTTTG 898  
Qt K L D V T N I K E A L E T F

RESULT 3















gene /gene="aq\_1347"  
CDS 939..1463  
/gene="aq\_1347"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07333.1"  
/db\_xref="PID:g2983774"  
/db\_xref="GI:2983774"  
/translation="MGFLERIKRLFSNEEKQESILSNPLERLKRKGLIETGETWE  
VTVALYDGASKEKEWEIRASRGFLSLEEGKIYFEEIDEPELEPDGEPHYRG  
KPEEYVYKGRKRLKYAGKAKYIKNLSESPVYIMPEFRSDGEMIDLEIMDEYENAY  
KGRLEDEWEESTLPR"  
1384..1953  
/gene="aq\_1348"  
CDS 1384..1953  
/gene="aq\_1348"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07334.1"  
/db\_xref="PID:g2983775"  
/db\_xref="GI:2983775"  
/translation="MRGWSGTRKLPKRGWYRNGKLRVEFLVSLLEILSCSSPKKVK  
GKPEEFLKNYEKGITRYALENKPQDTEYCKYHLKDLIKEGYIILSPSRIVEREK  
LSTRKLYCELSKGRFLVAGERTPPSALKVTVFELKGLFLMFLERVAAGGGGGG  
1950..2756  
/gene="aq\_1349"  
CDS 1950..2756  
/gene="aq\_1349"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07335.1"  
/db\_xref="PID:g2983776"  
/db\_xref="GI:2983776"  
/translation="MSGFLIALFWLFSKYVDFLFEAREKIREIFGNKNLALSISY  
AGYFLGLAFSYGVYSEFLREVLIIFVSPFLLLGVYIPDLIPRKIDLKKEILRG  
NAGAGITOGIYFLSILISAFMRKESFLSYISLYISLGMVLTSLTMSRL  
KNFEEVKKENSASLVISITLGVSVIYGAISGEPMGSLIFDLFSYLYIVVSQV  
LWYIFYVYVFLFRKYLISSEYVENNLSSILSNFLIASAFTITLAVMG"  
2753..4234  
/gene="aq\_1350"  
CDS 2753..4234  
/gene="aq\_1350"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC07337.1"  
/db\_xref="PID:g2983778"  
/db\_xref="GI:2983778"  
/translation="MSVSTRALKKICIFTFGAGIVAYSLATLATYLLGNVLYOMSVI  
STFLSMGUSHASRIIPDKPTPLAVLAEFLVLPSPVIAIHFANNEHLQTVI  
YGLSFVIGSLIGLEIPLAVRINNMVELKNNISVLEKDYLSVPAGLLYALFLPKL  
GFLPGLIILAGFENLISAFILVYKLRKFLKFAITFFLLATYAVGHRILYEEOK  
FGEELIIEEOPTPYOKIVLFRGKHSLYLDGHLQFSTDEKRYHETLVHVPSPFLK  
YKALILGGDGLALRELRKVPGEIHLVDLDKMEFSKKNLYMKRIKENSQVRL  
KVFSEDAFNPVKTKREKIDPYLDLDPRTPPSARARYSLEFYLSKNKLKEDGIFITQ  
AGDTFKREVFCISLTKTKAGFYAIPLVVYIPTFEGWGVIGSKPELNFENELEKEK  
TEFLNERALAEFTLTKSLSECPNVEVNTLKPVLIIYXXKION"  
4260..5306  
/note="aq\_1351"  
/gene="phoH"  
CDS 4260..5306  
/gene="phoH"  
/codon\_start=1  
/transl\_table=11  
/product="phosphate starvation-inducible protein"  
/protein\_id="AAC07324.1"  
/db\_xref="PID:g2983765"  
/db\_xref="GI:2983765"  
/translation="MEKVLAERIDIGRYDERFYAIVGNDENIKFFSKFVDVRIVPR

gene GHELIITKGEEDKVRVNFLLKAVLKRLEHERLTAERYEKAKOFLKSVEAKYKEQPG  
REKETLEEEVILITHRRKAIIVPTEYQKRYVENIKKNDIVFGIGPAGTKRTYLAAMA  
LQHLKDGKVNKIIILFRPAVEAGKLEFLGITAERKHPYLTPLPYDALYDMVYDKANY  
MLERNIIEIAPLAFMRGTINDAFLILDEAONATROMCMFLTRIIGSGSKAVTGGVT  
QIDLPKKESGILEAMEVLKNIEGIEFFIKQTDVVRHIVAKIINAYEQYEAKGOS  
SKTEPEESKEGOD"  
5230..5682  
/gene="aq\_1354"  
CDS 5230..5682  
/gene="aq\_1354"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC07338.1"  
/db\_xref="PID:g2983779"  
/db\_xref="GI:2983779"  
/translation="MSSTRKOKNRVLYKLRKRVKDKTEKNAELASLGNVNELS  
VITDDEIRBELNKTYRKDKPDVLSFPMGEFGYKILGDVVISODPAERQARELG  
HSLBEVYKRLIYGHVHLGLYDHEKGESEKFRLEENYVLSKSNAL"  
complement(5649..6479)  
/note="aq\_1355"  
/gene="lipA"  
CDS complement(5649..6479)  
/gene="lipA"  
/codon\_start=1  
/transl\_table=11  
/product="lipidic acid synthetase"  
/protein\_id="AAC07325.1"  
/db\_xref="PID:g2983766"  
/db\_xref="GI:2983766"  
/translation="MKKPVLFHSKLYEVKLLRKSHLYVCEESRCPNISCEFGNKTA  
TEMLGNCTRCACFCNVEKGFPGVGDVPLAEVATLGLKVVITISVTRDDLPD  
GGASHFAKCIIRVLENIEDIKYEVLPDFRGKKALAEKRPVLLNHNVEVPLRY  
PSVRIGANYSRLNITLMSKEIDKSYTSALILGGEKKEEYIKWEDLRSDVDPL  
VIGQYYPKSLKHHPVKYYSSEEFKFEEDIGYEMGRKFVSKPNRASSIKARESLIST  
"  
6505..6939  
/gene="aq\_1356"  
CDS 6505..6939  
/gene="aq\_1356"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07336.1"  
/db\_xref="PID:g2983777"  
/db\_xref="GI:2983777"  
/translation="MKGYIQFPPEPEKKEKELTEPLTEEDFNQVIEVKKDLT  
LTPSOVKDILSVIGEKYIGDERFVGLLMWERYKHSPLVSSFVAVLGDKLERISF  
LEAKNNYDMWKLMVITYKDKFTEELMRGLFELKREKEGEI"  
7056..8462  
/note="aq\_1357"  
/gene="cydA"  
CDS 7056..8462  
/gene="cydA"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome oxidase d subunit I"  
/protein\_id="AAC07328.1"  
/db\_xref="PID:g2983769"  
/db\_xref="GI:2983769"  
/translation="MDAVLIARIQPALTMAYHFLVPITLGLSLLAITIMTWAREED  
DERHILYERIGEFTLLFAINPAAGVAGIYLEFEGRWSEYSTVGIDFGAPLAIE  
ALFAFFLESVFGLVILFGDRDVSPEKFAESTMVAIGTSLAHLIIVANSWOOTPKG  
YIEGKAYIASFFDAVNSTPYRREYHNFALAVITTGATFVAGISAYIILKRHLVAK  
ESFRIAVTIGMTASLLQIILAGDLGVARITQPLKAMMEDWETESGGLTIVGIE  
IKPLSILAYHDPDAKYVGIKDLVOTFORIAGBEVGEDELPELMLDPROFAGRSYL  
ITEEELPTWIPFMSLRGMVFLGFFFAFTTSGFLTQDTTEORMLLKLFYSYPL  
PIVANKIWIIVAEVGROPVVYIILKTRMAASPLDAGQILTSIILFSSYIFVFLIFL  
GLMVQKVKRGRPREVA"  
8464..9564  
/note="aq\_1358"  
/gene="cydB"  
CDS 8464..9564



[illegible]

BASE COUNT	2787 a	750 c	957 g	2263 t
ORIGIN				
Query Match	2.3%	Score 31:	DB 18:	Length 6757:
Best Local Similarity	59.2%	Pred. No.	3.84e-02:	
Matches 100:	Conservative	0:	Mismatches 69:	Indels 0:
Gaps 0:				
/translation="MKRIGIIGALIEOEIRKKEILNNKIKKIGNIKRYGTEFKIEID LIISGIEKVSASASTTITSINLPDPDFLINSAGSINLACIDIDILIKPKTCYDVLD TNGVSGKQIPEYPOTEFKTNKRILREILKEIAVEFKFEPLUGLVTGDSFTRKSNCIK IKNOFSIAIGVDEMSSTAIGVCNFKIPIIILKISIDLSIDSNMATSHEKNIPIASKLS SKIVKLKLRKISSONSJ"				
Db	1171	TTAAAGATTATGACATCATCTACTGATTAATTAATCTGAAACATATAAACAATAGCAAAA	1230	
Qy	950	TAGAATTTATTTTGGACACCATCCAGAGAAATTAAGTCTACATTCACACGACAGCAAGA	1009	
Db	1231	GCTGSCCTAAAAAAATTTTAATTAATGATTTTCAACCTCATCAGATATACAGAACCCGAA	1290	
Qy	1010	AATATCCACATPAAGAAATTTGTGTGAGATTTTCAACCAACACACTTCTCTGACACAG	1069	
Db	1291	ATCATTAATTTGCATTTTATTAATAATTTTATCTCAAGTACACTTTATT	1339	
Qy	1070	CATTTTAAATGATTTGCAGAAAGTTATGTAAGACAGATCGTGATT	1118	
RESULT	11			
LOCUS	AE000736	15862 bp	DNA	BCF
DEFINITION	Aquifex aeolicus section 68 of 109 of the complete genome.			03-SEP-1998
ACCESSION	AE000736	AE000657		
NID	92983763			
VERSION	AE000736.1	GI:2983763		
KEYWORDS				
SOURCE	Aquifex aeolicus.			
ORGANISM	Aquifex aeolicus			
REFERENCE	Eubacteria; Aquificales; Aquificaceae; Aquifex.			
AUTHORS	1 (bases 1 to 15862) Decker,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L., Graham,D.E., Overbeek,R., Sneed,M.A., Keller,M., Aujay,M., Huber,R., Feldman,R.A., Short,J.M., Olson,G.J. and Swanson,R.V. The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.			
TITLE	Nature 392 (6674), 353-358 (1998)			
JOURNAL	98196666			
MEDLINE	2 (bases 1 to 15862)			
REFERENCE	Decker,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L., Graham,D.E., Overbeek,R., Sneed,M.A., Keller,M., Aujay,M., Huber,R., Feldman,R.A., Short,J.M., Olson,G.J. and Swanson,R.V.			
AUTHORS	Direct Submission Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego, CA 92121			
TITLE	CA 92121			
JOURNAL				
COMMENT	Putative indicates no similarity to known proteins Hypothetical indicates similarity to a protein of unknown function.			
FEATURES	Location/Qualifiers			
source	1..15862			
	/organism="Aquifex aeolicus"			
	/strain="VF5"			
	/db_xref="taxon:63363"			
	233..937			
	/gene="aq_1345"			
	233..937			
	/gene="aq_1345"			
	/codon_start=1			
	/transl_table=11			
	/product="putative protein"			
	/protein_id="AAC07332.1"			
	/db_xref="pid:92983773"			
	/db_xref="gi:2983773"			
	/translation="MMGLKRLILRVRAHSAVYEKLEDPKITEEGRLRKOLKEP MEALQVRAQSQIRFEAREARNKEERKALVKAESILLQAOEGKISPEAEKLAELIQL KHELEKARARLEEARAKOREKAVYLQAKIDELKVOIAKYAEAKTLKARATARA KNNKILAKVDPDTIAMLERMKEKEVEEETLAQAEELAKOVALDVESEKKEKLEKEV ESKOLTEPKRKLGLS"			
	939..1469			
gene				







TITLE  
JOURNALSubmitted (12-DEC-1997) The Institute for Genomic Research, 9712  
Medical Center Dr. Rockville, MD 20850, USALocation/Qualifiers  
1..15079FEATURES  
Source

## gene

## CDS

```
/organism="Borrelia burgdorferi"  
/db_xref="taxon:139"  
68..244  
/gene="BB0813"  
68..244  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon_start=1  
/transl_table=11  
/product="B. burgdorferi predicted coding region BB0813"  
/protein_id="AAC67176.1"  
/db_xref="PID:92688771"  
/db_xref="GI:2688771"  
/translation="MRGILYAILLYVFMVLMFCPAYFDITSTIFINIPLMVFLSGI  
LPPSIFEMVFLF"  
322..1656  
/gene="BB0814"  
322..1656  
/note="similar to PID:606198 GB:U00096 PID:1789656 percent  
identity: 33.64; identified by sequence similarity;  
putative"  
/codon_start=1  
/transl_table=11  
/product="pantothenate permease (panf)"  
/protein_id="AAC67168.1"  
/db_xref="PID:92688763"  
/db_xref="GI:2688763"  
/translation="MLNKYFLANRNINITYVALLFSSSYISASSRISGSPAYRYKGL  
STLLATIQITPTLVILIVGQRLNRESKINAINIIDYIRHRESDFLALSGEVL  
FSMFLISALIGGAKLIEVFGIDVVGLTFFAFVLYVFEGGKAVAYDLDLGG  
LMLVSVILFSLKMLDGGGINLFTKATSLDLSLPSNADLKPOYIISPMILGIG  
IIGPOIINNFIKDENAIKFSLPSTFSLISLYMLIGFPAILLEPDLSPNKY  
VUNAKVNPSPSCFMFYGLUSAINSTYDNLILTSYLISIRIYIKKDLKEDKYG  
RTIMSNIFFILILFSLPPEVLEFFINIFAGALEVSFFPIVFGTLNVS KIAA  
FASMFGLIFLYLILFELGINTWFFHFPSPFSFTFLVNVFCKKNSKVC"  
1666..2559  
/gene="BB0815"  
1666..2559  
/note="similar to GB:U002214 GB:I43967 SP:P47451 PID:406497  
PID:1045895 percent identity: 31.20; identified by  
sequence similarity; putative"  
/codon_start=1  
/transl_table=11  
/product="conserved hypothetical protein"  
/protein_id="AAC67167.1"  
/db_xref="PID:92688762"  
/db_xref="GI:2688762"  
/translation="MRDKYIFLEVANDNGKRLDILIKILNFKASIKIKIRKGD  
RNGKSHSPSCYKCGDKIYIKSLAONLITDQKFSKIDPOYIRKRIYEDSDIL  
VIDKQKIHGKNSLDFLVNSYLSQNLRSISFPSAVHLDRTSGIITFAKIN  
TKRUBEARSGSIIKRIYFALIGEVKSVVTKNLFKRLKTFVLDKNEVNAIT  
KNPILSCRAFLVEIVETGTHQIRSCSPNNHPLINDKRYCNKFKKSDYFLHAF  
LVNNGTFKKNKFCSPSLDF"  
2578..3528  
/gene="BB0816"  
2578..3528  
/note="hypothetical protein; identified by Glimmer;  
putative"
```

gene  
CDS

```
/codon_start=1  
/transl_table=11  
/product="B. burgdorferi predicted coding region BB0816"  
/protein_id="AAC67175.1"  
/db_xref="PID:92688770"  
/db_xref="GI:2688770"  
/translation="MMSNPFNFSSDFLKRILSKVKDPAISIKKHFVRVRYALVGS  
AGTGSFPRSHVADKYSLPLIDGVLKNNKIIIGSADFEDNYFKAVRSVFDDA  
HORELVEVLAKKEFNKILIGLSLKMDIKLIPVFKIYITDVTGQETEKAR  
ISRQGEHVPAARAEITSIRNLLNSIKIFPKSGWFPARKNIRSVRPHFEEG  
VLSIKRAVRQIIEHCVEIDNITVLDKIKKDGNNLEFLNIPILGNLNTNEM  
LNTYITVAKTIITNIIIDIVIKHFYKRDYLEESYG"  
3518..4924  
/gene="BB0817"  
3518..4924  
/note="similar to GB:I42023 SP:P45066 PID:1006469  
PID:1221261 PID:1205384 percent identity: 28.70;  
identified by sequence similarity; putative"  
/codon_start=1  
/transl_table=11  
/product="UDP-N-acetylmuramate--alanine ligase (murC)"  
/protein_id="AAC67166.1"  
/db_xref="PID:92688761"  
/db_xref="GI:2688761"  
/translation="MKVDFDNLNIFVYIGKSGACSLACPLNSKGYCEGVDSDF  
VTDELISNNKISYNDIYEFSLKODRSEFDLIVSSAANKGLOVLLFAKELNIPIS  
YPAELGELSKRTYSIGIGSHGKTTTFLVGLNFKLLNRYIVIGSSVKPKDMSAI  
AGISNIFVTECEYKRFHLPNSPNLLITNDYDEVDFKNEALEAFQYINLKL  
NGILINDSDNNLTKIKROIINKDLSISYSGDLSDFQISIAVSEYPCFSEGL  
NVELKTVEFHVHNSAALLALNLFLENSGKSIDPEEAIKRIAKNYSIGIRVEYK  
ENGVYVDD:AAHREIKNTLFGIKNFKNKRIITLDEMPHETFEKFEFADYVLS  
AADILILNIYISNRENPNDELVSYLELNKIKIKNRYFFKQVDSINFIKSLISG  
DLFTMGAGNNFILDPL"  
4935..5801  
/gene="BB0818"  
4935..5801  
/note="similar to GB:U10328 SP:P23839 GB:X14235 PID:290494  
PID:42721 percent identity: 29.96; identified by sequence  
similarity; putative"  
/codon_start=1  
/transl_table=11  
/product="conserved hypothetical protein"  
/protein_id="AAC67165.1"  
/db_xref="PID:92688760"  
/db_xref="GI:2688760"  
/translation="MKSMTGFFYELEIKIGNYFVNLKSYNGKLEFEKRLPEIFSGY  
DLIDNLSKYSIRGNVFLNKGKELVPSVNTIPNPIYEAISRDLASHTNLNKN  
ELSDGFLISGALIIDDESHOEIYGLFPGVLEALLHYNNGSPFGENTKSDIVS  
TIVLERDLKIKYKDCSDINVKLPASIKENISKLMDEPRDLINAEAKAKAIRDINE  
ETMRDHSIETPKYLEYICGKALEFISQEMHREITTSNKAVDLDIKLILNKKLN  
LEKIKQIRNVE"  
5798..6340  
/gene="BB0819"  
5798..6340  
/note="similar to GB:U77117 SP:Q58071 PID:1591369 percent  
identity: 38.76; identified by sequence similarity;  
putative"  
/codon_start=1  
/transl_table=11  
/product="cytidylate kinase (cmk-2)"  
/protein_id="AAC67164.1"  
/db_xref="PID:92688759"  
/db_xref="GI:2688759"  
/translation="MKIALSGSCGANTVSGMIKHYGLEFINYTFDIAREHNIP  
SEFYKEIIGRDYWDYLDNRSLVLSRKNNVTASRLAIWISADLKILYIAKME  
VRAERIMREGMYSDVLSSTFIRENKKRYLAIVNIDIDIDYFSETDLYIVDYNINP  
NEVPELIDEDIKRLKNS"  
6376..6576  
/gene="BB0820"  
6376..6576
```



QY 1276 CATATTCAAATTCACAAATGATATTAGATAATAGCATGAAATGCGTTTAA 1335  
Db 1811 WKMKRKAAMWTWR 1823  
QY 1336 TATGTTTATAATA 1348

RESULT 7  
LOCUS AF012089 10772 bp DNA INV 05-AUG-1997  
DEFINITION Drosophila melanogaster cysteine proteinase-1 (Cpi) gene, complete  
cds, and phenylalanyl tRNA synthetase gene, partial cds.  
ACCESSION AF012089  
MID g2305220  
VERSION AF012089.1 GI:2305220  
KEYWORDS fruit fly.  
SOURCE Drosophila melanogaster  
ORGANISM Eukaryote; mitochondrial eukaryotes; Metazoa; Arthropoda;  
Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 4546 to 4553)  
AUTHORS Gray, Y.H., Tanaka, M.M. and Sved, J.A.  
TITLE P-element-induced recombination in Drosophila melanogaster: hybrid  
element insertion  
JOURNAL Genetics 144 (4), 1601-1610 (1996)  
MEDLINE 97132596  
REFERENCE 2 (bases 1 to 10772)  
AUTHORS Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.  
TITLE Structure of the cysteine proteinase (Cpi) gene of Drosophila  
melanogaster and associated mutational effects  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 10772)  
AUTHORS Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.  
TITLE Direct Submission  
JOURNAL Submitted (30-JUN-1997) School of Biological Sciences, University  
of Sydney, Biology A12, Sydney University, NSW 2006, Australia  
FEATURES  
Source  
1..10772  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
join(872..1000,2310..2426,6476..6690,6751..7707)  
/gene="Cpi"  
/product="cysteine protease"  
872..7707  
/gene="Cpi"  
872..1000  
/gene="Cpi"  
/number=1  
1001..2309  
/gene="Cpi"  
2310..2426  
/gene="Cpi"  
/number=2  
join(2328..2426,6476..6690,6751..7462)  
/gene="Cpi"  
/codon\_start=1  
/product="cysteine proteinase-1"  
/protein\_id="AAB65749.1"  
/db\_xref="PIR:g2305221"  
/db\_xref="GI:2305221"  
translation="MRTAVLLPLALAAVAQVAFADVVEEMHTEKLEHKKRYODET  
EERFLKIFENKHKIAKNQRFKRAVKNKADLLHHEFRLMKNFYTLHK  
OLRAADESKGVTEISPAYTLPLPSVDMRTKAVTAKDGGSCAFSGTALGEO  
HERKSGVLSLSEONLVDCTKYGNGGNCGLMDNARYTKKNDGSGIDTEKSYSEAD  
DSCRNKGTGATDRGTDTDPQGDDEKMAEAVALTVGVSAIDAHSFQFSEGYTN  
EPQCDQNDLHGVLVVGFGEDESGEDYLVKNSGTTWGDKGFKMLRNKNGCGIAS  
ASSYPLV"  
2427..6475  
/gene="Cpi"  
4546..4553  
/gene="Cpi"  
/note="insertion site of P[Casper](50C)"  
/citation=[1]

exon 6476..6690  
/gene="Cpi"  
/number=3  
6691..6750  
/gene="Cpi"  
6751..7707  
/gene="Cpi"  
/number=4  
join(<8110..9300,9370..>9532)  
/product="phenylalanyl tRNA synthetase"  
join(8110..9300,9370..>9532)  
/note="potential orf"  
/codon\_start=1  
/product="phenylalanyl tRNA synthetase"  
/protein\_id="AAB65750.1"  
/db\_xref="PIR:g2305222"  
/db\_xref="GI:2305222"  
translation="MLITLTVQCARHKLKSTRCLASSAPAKSPSPPOLEVGSTIYA  
TDCMTNTPKILSYGANKHLDTHDPLUSTIRORIVNFTGATRNQKNPLESVYDQNN  
PVYTVQNEPMLIPADHVSRSKSDCYIIHQHLRAHTAHQVELISGLDNLVVG  
EYVRDEIDSTHYVPHQADAVRLVTKDKLFERNPGLEFEETSGSLADPKLHP  
SSWTKRSPATRRARVAKMEHMKHVLGTLKDLFGPRIKRWVDTEPEPQSPMELEI  
YKDNMLEVYGCGIMHEILORSGVHOSIGARGVGERLAWLPIDPIRLFMSNDS  
GRFSQSEKDLNLPKPKPISHIPQCTNDLSFPLPDIEVDAGFSRDPFDLYRSVAG  
DVEQQLSLVDKFRHFKTKSSVCFRIYRMRMTLTQAEVNEITHKQIASVDSFNVQ  
IX"

BASE COUNT 2929 a 2357 c 2282 g 3046 t 158 others  
ORIGIN

Query Match 2.8%; Score 38; DB 21; Length 10772;  
Best Local Similarity 15.6%; Pred. No. 8.76e-06;  
Matches 23; Conservative 72; Mismatches 51; Indels 1; Gaps 1;

Db 1683 YTYTKTAARKRWYMAWTTTMMKMMWTRTMMKMMKTYRTMMKMMWTRTTTMMW 1742  
Cp 594 TTCTTGATATCATCAAAACATCATTAATATCTTTGAATAATAC-AGAGATGCAAT 536  
Db 1743 YTWSTWTKRYMAYMMKMMWTRTMMKMMKRTSAAYASAMRKMMKMMWYRAM 1802  
Cp 535 CAATATTTGTCATATATGCGTAATCAGGTTATTAAGCTAAAGTACGCTATATTCAC 476  
Db 1803 KTYMMAAMKMKRKAAMWTWRMYMTT 1829  
Cp 475 ATGCCTCAAAAGCGAATAATCACTT 449

RESULT 8  
LOCUS AE001180 15079 bp DNA BCT 15-DEC-1997  
DEFINITION Borrelia burgdorferi (section 66 of 70) of the complete genome.  
ACCESSION AE001180 AE000783  
MID g2688755  
VERSION AE001180.1 GI:2688755  
KEYWORDS Lyme disease spirochete.  
SOURCE Borrelia burgdorferi  
ORGANISM Eubacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia  
burgdorferi group.  
1 (bases 1 to 15079)  
Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A.,  
Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K.,  
Galin, M., Dougherty, B., Tomb, J.-F., Fleischmann, R.D.,  
Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J.,  
Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M.D.,  
Gocayne, J.D., Weiman, J., Ufferback, T., Matthey, L., McDonald, L.,  
Artlich, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K.,  
Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C.  
TITLE Genomic sequence of a Lyme disease spirochete, Borrelia  
burgdorferi  
JOURNAL Nature 390 (6660), 580-586 (1997)  
MEDLINE 98065943  
REFERENCE 2 (bases 1 to 15079)  
AUTHORS Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A.,  
Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K.,



(TM)

Distribution rights by Oxford Molecular Ltd

n.a. - n.a. Smith-Waterman search, using a protein database which has been backtranslated into n.a. using IUPAC symbols

Wed Nov 24 00:54:39 1999;

## TACTICATTC

TABLE bktranslated

Dbase 0; Query 0

170751 seqs, 63799824 bases x 2

Minimum Match 08  
Listing first 45 summaries

a-geneseq35

1.pat1 2.pat2 3.pat3 4.pat4 5.pat5 6.pat6 7.pat7  
8.pat8 9.pat9 10.pat10 11.pat11 12.pat12 13.pat13  
14.pat14 15.pat15 16.pat16 17.pat17 18.pat18  
19.pat19 20.pat20 21.pat21 22.pat22 23.pat23  
24.pat24 25.pat25 26.pat26 27.pat27 28.pat28  
29.pat29 30.pat30 31.pat31 32.pat32 33.pat33  
34.pat34 35.pat35 36.pat36 37.pat37 38.pat38  
39.pat39

Mean 58.676; Variance 331.404; scale 0.177

**Pred.** No. Is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	5363	79.4	437	39	W877771	UDP-N-acetylmutamate:	0.00e+00
2	2637	39.0	215	39	W891999	Partial sequence of t	7.63e-17
3	1730	25.6	422	33	W55120	Streptococcus pneumoniae	8.14e-11
4	544	8.1	46	34	W75686	Staphylococcus aureus	2.63e-24
5	182	2.7	335	33	W55117	Streptococcus pneumoniae	7.75e-01
6	180	2.7	395	38	W03474	Mouse SR $\alpha$ -related pro	9.99e-01
7	182	2.7	450	37	W68551	S. pneumoniae MurD pr	7.75e-01
8	182	2.7	450	26	W29454	Streptococcus pneumoniae	7.75e-01
9	178	2.6	286	22	W20102	H. pylori cytoplasmic	1.29e+00
10	178	2.6	286	22	W24505	H. pylori cytoplasmic	1.29e+00
11	178	2.6	455	22	W20606	H. pylori cytoplasmic	1.29e+00
12	175	2.6	1529	18	R97985	CORK potasssium channel	1.88e+00

[illegible]



CC drug design and as an immunogen for vaccines. The Murt gene sequences are  
CC useful in antisense/xrbozyme therapeutics; to detect mutant Murt gene;  
CC for chromosomal mapping; to determine bacterial serotype; and for genetic  
CC immunisation.  
S0 Sequence 437 AA;

Query Match	Score	DB	Length
79.48;	5363;	39;	437;

Best Local Similarity 60.9%; Pred. No. 0.00e+00;  
Matches 799; Conservative 320; Mismatches 192; Indels 0; Gaps 0;

Dh	1	atganc	eyt	ac	tyt	cyt	lty	gt	ngn	ath	aa	rgn	sm	ngn	aa	rgn	sm	ny	ngn	cn	ar	h	60		
Dt	22	atgac	ac	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	81		
Qy	82	atgac	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	81	
Qt		m	t	h	y	h	f	v	g	i	k	g	s	g	m	s	s	l	a	q	i				
Dh	61	atgca	g	ay	u	ngn	cn	ca	g	ay	u	ngn	cn	ca	g	ay	u	ngn	cn	ca	g	ay	u	ngn	120
Dt	142	gtt	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	141		
Qy	82	atgca	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	141	
Qt		m	h	d	l	g	h	e	v	o	g	s	d	i	e	n	y	v	f	t	e				
Dh	121	g	u	ngn	cn	u	ngn	cn	u	ngn	cn	u	ngn	cn	u	ngn	cn	u	ngn	cn	u	ngn	cn	180	
Dt	142	gtt	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	141		
Qy	142	gtt	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	ct	141		
Qt		v	a	l	r	n	k	g	i	k	i	l	p	f	g	a	n	i	k	e					
Dh	181	g	ay	u	ngn	cn	u	ngn	cn	u	ngn	cn	u	ngn	cn	u	ngn	cn	u	ngn	cn	u	ngn	240	
Dt	202	g	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	261	
Qy	202	g	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	g	at	261	
Qt		d	m	v	i	q	g	n	a	f	a	s	s	h	e	i	v	r	a						
Dh	241	h	q	i	k	i	d	v	v	s	y	n	d	f	i	g	q	i	d	q					
Dt	262	h	q	i	k	i	d	v	v	s	y	n	d	f	i	g	q	i	d	q					
Qy	262	h	q	i	k	i	d	v	v	s	y	n	d	f	i	g	q	i	d	q					
Qt		h	q	i	k	i	d	v	v	s	y	n	d	f	i	g	q	i	d	q					
Dh	301	y	t	s	v	a	v	t	g	a	h	g	k	t	s	t	g	l	s						
Dt	322	y	t	s	v	a	v	t	g	a	h	g	k	t	s	t	g	l	s						
Qy	322	y	t	s	v	a	v	t	g	a	h	g	k	t	s	t	g	l	s						
Qt		y	t	s	v	a	v	t	g	a	h	g	k	t	s	t	g	l	s						
Dh	361	h	v	m	n	g	d	k	k	t	s	f	i	g	d	g	t	g	m	g					
Dt	382	h	v	m	n	g	d	k	k	t	s	f	i	g	d	g	t	g	m	g					
Qy	382	h	v	m	n	g	d	k	k	t	s	f	i	g	d	g	t	g	m	g					
Qt		h	v	m	n	g	d	k	k	t	s	f	i	g	d	g	t	g	m	g					
Dh	421	l	p	e	s	d	y	f	a	f	e	a	c	e	y	r	h	f	l	s					
Dt	442	l</																							

[illegible]







Query Match 25.6%; Score 1730; DB 33; Length 422;  
 Best Local Similarity 42.7%; Pred. No. 8,14e-111;  
 Matches 525; Conservative 218; Mismatches 472; Indels 14; Gaps 12;

```

Db      h q m g h k v q g s d v e k y y f t q r
Dc      1 cayaraatgggncayaarglncargmsngayltngaraarlaytlaycnacmgn 60
QY      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      85 CATATTAGTACAGATCGATTCAGATCGATTCAGATTCAGATTCAGATTCAGATTC 144
Qc      H D L G H E V Q G S D I E N Y V F T E V

Db      g l e q a g i t i l p f d e k n l d g d
Dc      61 ggnylngarcrgcnngnathcnathcnathcnathcnathcnathcnathcnath 120
QY      145 GCTGTAGTACAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 204
Qc      A L R N K G I K I L P F G A N N I K E D

Db      m e i l a g n a f r p d n n v e i a y a
Dc      121 atgaratthahgcnngnagncnngnagncnngnagncnngnagncnngnagncn 180
QY      205 ATGTAGTACAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 261
Qc      M V V I O G N A F X R X S H E E I V R A

Db      d q n g i s y k r y h e f l g s f m r d
Dc      181 gayaraaaygnathwsntayaa-rmgntaycayartlytngnswntlytngmng 239
QY      262 CATCAAT-TGAATTAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGAT 320
Qc      H Q X E I R C C S Y N D F L G Q I I D Q

Db      f v s m g v a g a h g k t s t g m l s
Dc      240 yltglnwsnaltgngngngngngngngngngngngngngngngngngngngng 299
QY      321 ATTAAGTACAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 380
Qc      Y T S V A V T G A H G K T S T T G L L S

Db      h v l s h i t d t s f l i g d g t g r g
Dc      300 ncaygltynsnacayathacnagaycnwntlytlnathgngaygnagngmng 359
QY      381 ACAGTGTACAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 440
Qc      H V M N G D K K K T S F L I G D G T G M G

Db      s a n a k y f v f e s d e y e r h f m p
Dc      360 nwsngncaaycnarlaytlytngaytngaytngaytngaytngaytngaytng 419
QY      441 ATTCAGTACAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 500
Qc      L P E S D Y F A F E A C E Y R H F L S

Db      y h p e y s i l t n i d f d h p d y f t
Dc      420 ntaycaycngarlayawnathacnathacnathacnathacnathacnathac 479
QY      501 TTATAACCTATTAAGCATTAATTAAGCATTAATTAAGCATTAATTAAGCATTA 560
Qc      Y K P D Y A I M T N I D F D H P D Y F K

Db      s l e d v f n a f n d y a k q i t k g l
Dc      480 nwsytlngarlaytlytngaytngaytngaytngaytngaytngaytngay 539
QY      561 AGATATTAAATGATGTTTGTATTCATTCAGATTCAGATTCAGATTCAGATTC 620
Qc      D I N D V F D A F Q E M A H N V K K G I

Db      f v y g e d a e l r k i t s d a p l y y
Dc      540 ntlytntlayngarlaytngaytngaytngaytngaytngaytngaytngay 599
QY      621 TATGCTTGCGGTATGATTAATTAAGCATTAATTAAGCATTAATTAAGCATTA 680
Qc      I A W G D D E H L R K I E A D V P I Y Y

Db      y g f e a e g n d f v a s d l l r s i t
Dc      600 ytagygtlytngarngaytngaytngaytngaytngaytngaytngaytngay 659
QY      681 CTATGATTTTAAAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGAT 737
Qc      Y G F K D S D D X L X A Q N I Q I T D K

Db      g s t f t v h f r g q n l g q f h l p t
Dc      660 ngwnsnacntlycngntncaytlymngngncaraaytngngncarlaycaythcnac 719

```

```

QY      738 AGGATCTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 797
Qc      G T A F D V Y V D G E F Y D H F L S P Q

Db      f g r h n i m n a t a v i g l i y t a g
Dc      720 ntlygmngncayaayathatgaaycnngnathgnytnytnytnytnytnytny 779
QY      798 ATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 857
Qc      Y G D H T V L N A L A V I A I S Y L E K

Db      f d l n l v r e h l k t f a g v k r r f
Dc      780 ntlygaytlnaytngtngngarlaytngaytngaytngaytngaytngaytng 839
QY      858 GCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 917
Qc      L D V T N I K E A L E T F G G V K R F

Db      t e k i v n d t v l i d d f a h h p t e
Dc      840 ycnagaraathgtnaayacngnathhahgayaytlycncayaaycnacng 899
QY      918 CAATGAACATCAATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 977
Qc      N E T T I A N Q V I V D Y A H H P R E

Db      i a t l d a a r q k y p s k e i v a v
Dc      900 rathahgcnaytngaycngngngngngngngngngngngngngngngngngng 959
QY      978 AATTAGTGTACATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 1037
Qc      I S A T I D T A R K K Y P H K E V V A V

Db      f q p h t f t r t l a l l d d f a h a l
Dc      960 ntlycarncaayacntlycngnathcnathcnathcnathcnathcnathcnath 1019
QY      1038 ATTCAACACACACATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1097
Qc      F Q P H T F S R T Q A F L N E A F A E S L

Db      n q a d a v y l a c q i y g s a r e v d h
Dc      1020 naaycargngaycngntnaytngncarhathgngnswngngngngngngng 1079
QY      1098 ATGTAACGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 1154
Qc      C K A D R V F L C E I F G S I R E N X X

Db      g d v k v e d l a n k l n k k h q v i t
Dc      1080 ysgngaytngaytngaytngaytngaytngaytngaytngaytngaytng 1139
QY      1155 TGCGCATTAACATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1211
Qc      G A L T I Q D L I D K I G X V H X F I N

Db      v e n v s p l l d h d n a v y v f m g a
Dc      1140 ngtnaraaytngnncnnytnytnytnytnytnytnytnytnytnytnytn 1199
QY      1212 TGAAGATCTTATTAAGTATTAAGATTCATTCATTCATTCATTCATTCATTC 1271
Qc      E D L I N V L E Q F D N A V V L F M G A

Db      g d i q t y e y s
Dc      1200 ngngayathcarcnlaygarlaywsnt 1228
QY      1272 AGGATATTCATAAATTAACAAATGAT 1300
Qc      G D I Q K L Q N A

Db      4
ID      777686 standard; Protein; 46 AA.
AC      777686;
DE      30-OCT-1998 (first entry)
DE      Staphylococcus aureus protein of unknown function.
KW      Staphylococcus aureus protein; immune response induction; eye infection;
KW      antibody production; T-cell immune response; gastrointestinal infection;
KW      respiratory infection; inhibitor; bacterial infection; cardiac infection;
KW      central nervous system; kidney infection; urinary tract infection;
KW      antimicrobial compound identification; broad spectrum antibiotic;
KW      therapy.
OS      Staphylococcus aureus.
PN      EP-841394-A2.
PD      13-MAY-1998.

```



















RESULT 14  
ID R2282 standard; Peptide; 26 AA.  
AC R2282;  
DE 22-AUG-1992 (first entry)  
DE Amphiphilic peptide to inhibit growth of a target cell.  
KW Antimicrobial; antifungal; antitumor; pharmaceuticals; antibiotics;  
KW Antiparasitic; spermicides; burns; wound healing.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT modified\_site 1  
FT modified\_site /note= "acylated"  
FT modified\_site /note= "amidated"  
PN CA2047317-A.  
PD 20-JAN-1992.  
PE 18-JUL-1991; 024317.  
PR 19-JUL-1990; US-554442.  
PR 08-JUL-1991; US-725331.  
PA (SCRI-) SCRIPPS RES INST.  
PI Houghten RA, Blondelle S;  
DR WPI; 92-114943/15.  
PT Amphiphilic peptide(s) and analogues - for use in e.g.  
PT antimicrobial, antifungal or antitumor compositions, having  
PT increased biological activity  
PS Claim 30; Page 60; 71pp; English.  
CC The amphiphilic peptide (SEQ ID NO 61) was prep'd. by standard  
CC solid phase methods. The peptide is an analogue of the peptide  
CC having SEQ ID NO 3 (R22824), with an N-terminal extension of 8  
CC amino acids. Substitution and deletion analogues of this peptide  
CC have increased biological activity and are effective as pharma-  
CC ceuticals e.g. antibiotics for bacterial, fungal or viral  
CC infections, or in spermicides or antitumor or antiparasitic  
CC agents. Additionally the peptides can be used in wound healing  
CC comps. or for treating burns or other skin or eye infections.  
CC See also R2282-89.  
SQ Sequence 26 AA;

Query Match 2.3%; Score 155; DB 4; Length 26;  
Best Local Similarity 34.9%; Pred. No. 2.18e+01;  
Matches 22; Conservative 20; Mismatches 21; Indels 0; Gaps 0;

DB k l l k k l l k l l k k l l k k l  
Dt 6 raarytnynaaraarytnynaaraarytnynaaraarytnynaaraaryt 65  
OY 559 AAGAT 618  
Qt K I L M M F L M H S K K W H I M L K K V

DB 66 nyl 68  
Dt :l  
OY 619 ATT 621  
Qt

RESULT 15  
ID P91337 standard; peptide; 28 AA  
AC P91337  
DE 19-MAR-1990 (first entry)  
DE Amino acid sequence of Shiva-5.  
KW Shiva-5; lytic peptide; antimicrobial peptide; disease-resistant  
KW trichophyte; Shiva-2; Shiva-3; Shiva-4; Shiva-6; Shiva-7.  
PN W089004371-A.  
PD 18-MAY-1989.  
PE 2-NOV-1988; U03908.  
PR 02-NOV-1987; US-115941.  
PA (LODU) Louisiana State Univ.  
PI Jaynes JM, Derrick KS;  
DR WPI; 89-165650/22.  
PT Transformed plants contg. heterologous gene - expressing antimicrobial  
PS Table I; 56pp; English.  
CC agent, or polypeptide high in essential amino acids  
CC Amino acid sequence of Shiva-5 as an exemplary lytic peptide for  
CC use as an antimicrobial peptide contemplated for use in plant

CC (trichophyte) transformants in the invention. It is a homologue of  
CC Shivas 1-4, 6 and 7. All of these Shiva peptides are also contemplated as  
CC having general utility in inducing lysis of cells in vitro.  
SQ Sequence 28 AA;

Query Match 2.3%; Score 155; DB 1; Length 28;  
Best Local Similarity 32.9%; Pred. No. 2.18e+01;  
Matches 24; Conservative 21; Mismatches 28; Indels 0; Gaps 0;

DB l k l l k k l l k k l l k k l l k k  
Dt 11 tnytnaarytnynaaraarytnynaaraarytnynaaraarytnynaara 70  
OY 555 TTCAAGAT 614  
Qt S K I L M M F L M H S K K W H I M L K K

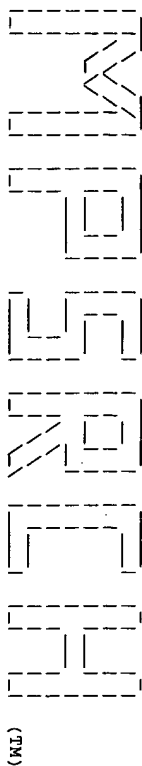
DB l l k  
Dt 71 aarytnynaaryt 83  
OY 615 AGGTAT 627  
Qt V L L

Search completed: Wed Nov 24 01:06:50 1999  
Job time : 731 secs.



**THIS PAGE BLANK (USPTO)**





(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_ntp n.a. - n.a. Smith-Waterman search, using a protein database  
which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Nov 24 01:07:37 1999; Maspar time 75.15 Seconds

Tabular output not generated. 1285.007 Million cell updates/sec

Title: >US-09-103-287-1  
Description: (1-1351) from US09103287.seq

Perfect Score: 6755  
N.A. Sequence: 1 ATGAGTAGGAGCTTTATAT.....TTAATATGTTTATATAGAG 1351

Comp: TACATATTCCTCAATATATA.....AATATACAAATATATATC

Scoring table: TABLE bkttranslate2  
Gap 30

Mmatch STD : Dbase 0; Query 0

Searched: 122461 seqs, 35738955 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfile1

Statistics: Mean 55.325; Variance 346.966; scale 0.159

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	182	2.7	450	2	US-08-843-Sequence 2, Applicatio	7.82e+01
2	182	2.7	450	1	US-08-665-Sequence 2, Applicatio	7.82e+01
3	177	2.6	267	2	US-08-934-Sequence 4, Applicatio	1.41e+00
4	177	2.6	446	2	US-08-934-Sequence 2, Applicatio	1.41e+00
5	155	2.3	26	1	US-07-725-Sequence 61, Applicati	1.75e+01
6	155	2.3	26	3	PCT-US91-0 Sequence 62, Applicati	1.75e+01
7	155	2.3	30	1	US-07-725-Sequence 62, Applicati	1.75e+01
8	155	2.3	30	3	PCT-US91-0 Sequence 62, Applicati	1.75e+01
9	155	2.3	36	1	US-07-725-Sequence 63, Applicati	1.75e+01
10	155	2.3	36	3	PCT-US91-0 Sequence 63, Applicati	1.75e+01
11	148	2.2	60	1	US-08-117-Sequence 20, Applicati	3.79e+01
12	147	2.2	225	1	US-07-991-Sequence 3, Applicatio	4.23e+01
13	147	2.2	225	1	US-08-107-Sequence 3, Applicatio	4.23e+01
14	147	2.2	225	1	US-08-544-Sequence 3, Applicatio	4.23e+01
15	148	2.2	416	1	US-08-117-Sequence 62, Applicatio	3.79e+01
16	149	2.2	416	1	US-08-117-Sequence 62, Applicatio	3.40e+01
17	150	2.2	3169	2	US-08-477-Sequence 6, Applicati	3.05e+01
18	143	2.1	442	4	5310667-6 Patent No. 5310667.	6.53e+01
19	143	2.1	444	1	US-08-476-Sequence 51, Applicati	6.53e+01

20	143	2.1	444	2	US-08-833-Sequence 51, Applicati	6.53e+01
21	143	2.1	444	1	US-08-306-Sequence 51, Applicati	6.53e+01
22	140	2.1	552	1	US-07-999-Sequence 22, Applicati	9.01e+01
23	140	2.1	552	1	US-07-999-Sequence 24, Applicati	9.01e+01
24	140	2.1	552	1	US-08-426-Sequence 22, Applicati	9.01e+01
25	140	2.1	552	1	US-08-401-Sequence 22, Applicati	9.01e+01
26	140	2.1	552	1	US-08-401-Sequence 24, Applicati	9.01e+01
27	140	2.1	552	1	US-08-426-Sequence 24, Applicati	9.01e+01
28	139	2.1	2109	3	PCT-US96-0 Sequence 6, Applicatio	1.00e+02
29	135	2.0	22	1	US-07-725-Sequence 60, Applicati	1.53e+02
30	135	2.0	22	3	PCT-US91-0 Sequence 60, Applicati	1.53e+02
31	137	2.0	132	2	US-08-649-Sequence 86, Applicati	1.24e+02
32	137	2.0	132	2	US-08-649-Sequence 97, Applicati	1.24e+02
33	137	2.0	132	2	US-08-649-Sequence 98, Applicati	1.24e+02
34	137	2.0	132	2	US-08-649-Sequence 108, Applicat	1.24e+02
35	137	2.0	132	2	US-08-649-Sequence 62, Applicati	1.38e+02
36	136	2.0	132	2	US-08-649-Sequence 88, Applicati	1.38e+02
37	137	2.0	135	2	US-08-649-Sequence 109, Applicati	1.24e+02
38	136	2.0	253	2	US-08-720-Sequence 6, Applicatio	1.38e+02
39	135	2.0	852	2	US-08-463-Sequence 36, Applicati	1.53e+02
40	135	2.0	884	2	US-08-471-Sequence 5, Applicatio	1.53e+02
41	135	2.0	884	2	US-08-471-Sequence 5, Applicatio	1.53e+02
42	137	2.0	900	2	US-09-005-Sequence 62, Applicati	1.24e+02
43	137	2.0	900	2	US-08-630-Sequence 62, Applicati	1.24e+02
44	135	2.0	1338	2	US-08-463-Sequence 50, Applicati	1.53e+02
45	135	2.0	1346	2	US-08-470-Sequence 23, Applicati	1.53e+02

## ALIGNMENTS

RESULT ID	1	US-08-843-309-2	STANDARD:	PRT:	450 AA.
AC	xxxxxx				
DE	Sequence 2, Application US/08843309				
CC	Sequence 2, Application US/08843309				
CC	Patent No. 5834270				
CC	GENERAL INFORMATION:				
CC	APPLICANT: Skatrud, Paul				
CC	APPLICANT: Peery, Robert				
CC	APPLICANT: Hoskins, Joann				
CC	APPLICANT: No, Chyun-Tea Earnest				
CC	TITLE OF INVENTION: Biosynthetic Gene Mur D of Streptococcus pneumoniae				
CC	NUMBER OF SEQUENCES: 3				
CC	CORRESPONDENCE ADDRESS:				
CC	ADDRESSEE: Eli Lilly and Company				
CC	STREET: Lilly Corporate Center				
CC	CITY: Indianapolis				
CC	STATE: Indiana				
CC	COUNTRY: US				
CC	ZIP: 46285				
CC	COMPUTER READABLE FORM:				
CC	MEDIUM TYPE: Floppy disk				
CC	COMPUTER: IBM PC compatible				
CC	OPERATING SYSTEM: PC-DOS/MS-DOS				
CC	SOFTWARE: Patent Release #1.0, Version #1.30				
CC	CURRENT APPLICATION DATA:				
CC	APPLICATION NUMBER: US/08/843.309				
CC	FILING DATE:				
CC	CLASSIFICATION: 435				
CC	ATTORNEY/AGENT INFORMATION:				
CC	NAME: Webster, Thomas D				
CC	REGISTRATION NUMBER: 39, 872				
CC	REFERENCE/DOCKET NUMBER: X-9900				
CC	TELECOMMUNICATION INFORMATION:				
CC	TELEPHONE: 317-276-3334				
CC	TELEFAX: 317-276-3861				
CC	INFORMATION FOR SEQ ID NO: 2:				
CC	SEQUENCE CHARACTERISTICS:				



CC LENGTH: 450 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 450 AA: 48579 MW: 1010380 CN:

Query Match 2.7%; Score 182; DB 2; Length 450;  
Best Local Similarity 41.8%; Pred. No. 7.82e-01;  
Matches 46; Conservative 18; Mismatches 46; Indels 0; Gaps 0;

DB H N V E N A L A T I A V A K L R D V D N  
Dt 832 CAYAAVGTNGARAAVGCNTNGCNACNATHGCGNGTNGCAARTTNGCAVGTNGAYAY 891  
Qy 808 CATACAGTTTAAATGCAATTAAGCTTAATTCGATTAGTATTAGAGAGCTAGATGTT 867  
Qt H T V L N A L A V I A I S Y L E K L D V

DB O T I K E T L S A F G G V K H R  
Dt 892 CARACATTAAGARACNTTNSGCTTYGNGGNGTNAARCAVNGNT 941  
Qy 868 ACAAAATATTAAGAACATTAAGAACGTTGGTGTGTTAAACGTCGTT 917  
Qt T N I K E A L E T F G G V K R R

RESULT 2  
ID US-08-665-435A-2 STANDARD: PRT; 450 AA.

AC xxxxxx

Sequence 2, Application US/08665435A  
Patent No. 5681694

GENERAL INFORMATION:

APPLICANT: Skatrud, Paul  
APPLICANT: Peery, Robert  
APPLICANT: Hoskins, Joann  
APPLICANT: Wu, Chyun-Yeh Earnest  
TITLE OF INVENTION: Biosynthetic Gene Mord of Streptococcus  
TITLE OF INVENTION: pneumoniae  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: US

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/665,435A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Webster, Thomas D

REGISTRATION NUMBER: 39,872

REFERENCE/DOCKET NUMBER: X-9900

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-3334

TELEFAX: 317-276-3861

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 450 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 450 AA: 48579 MW: 1010380 CN:

Query Match 2.7%; Score 182; DB 1; Length 450;  
Best Local Similarity 41.8%; Pred. No. 7.82e-01;  
Matches 46; Conservative 18; Mismatches 46; Indels 0; Gaps 0;

DB H N V E N A L A T I A V A K L R D V D N  
Dt 832 CAYAAVGTNGARAAVGCNTNGCNACNATHGCGNGTNGCAARTTNGCAVGTNGAYAY 891  
Qy 808 CATACAGTTTAAATGCAATTAAGCTTAATTCGATTAGTATTAGAGAGCTAGATGTT 867  
Qt H T V L N A L A V I A I S Y L E K L D V

DB O T I K E T L S A F G G V K H R  
Dt 892 CARACATTAAGARACNTTNSGCTTYGNGGNGTNAARCAVNGNT 941  
Qy 868 ACAAAATATTAAGAACATTAAGAACGTTGGTGTGTTAAACGTCGTT 917  
Qt T N I K E A L E T F G G V K R R

RESULT 3  
ID US-08-934-481-4 STANDARD: PRT; 267 AA.

AC xxxxxx

Sequence 4, Application US/08934481

Patent No. 5929045

GENERAL INFORMATION:

APPLICANT: Wallis, Nicholas G.  
APPLICANT: Fueyo, Joanna L.  
APPLICANT: Lonetto, Michael A.  
TITLE OF INVENTION: NOVEL MURD  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/934,481

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Dickinson, Todd Q

REGISTRATION NUMBER: 28,354

REFERENCE/DOCKET NUMBER: GM10070A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-994-2252

TELEFAX: 215-994-2222

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 267 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE 267 AA: 29079 MW: 348410 CN:

Query Match 2.6%; Score 177; DB 2; Length 267;  
Best Local Similarity 40.9%; Pred. No. 1.41e+00;  
Matches 45; Conservative 18; Mismatches 47; Indels 0; Gaps 0;



Dbb K L T L K R L K L T K R L K R L  
6 RAAATTTTAAARAKLTNTTNARTNTNAAAARATTTTAAARAKLTAARART 65  
Dc :  
Oy AAGATTATATGATGTGTTTTGCATTCACGAAGAATGCACATATGTTAAATAAAGT 618











Db 96 NYT 98  
Dt :  
OY 619 ATT 621  
Qt

RESULT 10  
ID PCT-US91-05047-63 STANDARD: PRT: 36 AA.  
XX AC xxxxxx  
XX

Sequence 63, Application PC/TUS9105047

Sequence 63, Application PC/TUS9105047

GENERAL INFORMATION:

APPLICANT: Houghten, Richard

APPLICANT: Blondelle, Sylvie

TITLE OF INVENTION: Amphiphilic Peptide Compositions and

TITLE OF INVENTION: Analogues Thereof

NUMBER OF SEQUENCES: 68

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Dressler, Goldsmith, Suter, Shore,

ADDRESSEE: 6 Milnamow

STREET: 180 North Stetson

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US91/05047

FILING DATE: 19910717

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/554,422

FILING DATE: 19-JUL-1990

ATTORNEY/AGENT INFORMATION:

NAME: Gamson, Edward P.

REGISTRATION NUMBER: 29,381

REFERENCE/DOCKET NUMBER: 421250-80

TELECOMMUNICATION INFORMATION:

TELEPHONE: 3126165418

TELEFAX: 3126165460

INFORMATION FOR SEQ ID NO: 63:

SEQUENCE CHARACTERISTICS:

LENGTH: 36 amino acids

TYPE: AMINO ACID

STRANDEDNESS:

MOLECULE TYPE: linear

FEATURE:

OTHER INFORMATION: acetylated at N-terminus, may be

OTHER INFORMATION: a C-terminal amide

SEQUENCE 36 AA: 4362 MW: 7668 CN:

Query Match 2.3%: Score 155; DB 3; Length 36;  
Best Local Similarity 34.9%; Pred. No. 1.75e+01;  
Matches 22; Conservative 20; Mismatches 21; Indels 0; Gaps 0;

Db K L L K K L L K L K L K L K L K L  
Dt : : : : : : : : : : : : : : : : : :  
OY 559 AAGGATATTAAGATGTTTTCATTCGACGAAGATGATTAAGTAAAGGT 618  
Qt K I L M M F L M H S K K W H I M L K K Y

Db 96 NYT 98  
Dt :  
OY 619 ATT 621  
Qt

RESULT 11  
ID US-08-117-083-20 STANDARD: PRT: 60 AA.  
XX AC xxxxxx  
XX

Sequence 20, Application US/08117083

Sequence 20, Application US/08117083

Patent No. 571954

GENERAL INFORMATION:

APPLICANT: Boursnell, Michael E.

APPLICANT: Inglis, Stephen C.

APPLICANT: Munro, Alan J.

TITLE OF INVENTION: Recombinant Virus Vectors Encoding Human

TITLE OF INVENTION: Papilloma Virus Proteins

NUMBER OF SEQUENCES: 70

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Walter H. Dreger

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/117,083

FILING DATE: 10-SEP-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Dreger, Walter H.

REGISTRATION NUMBER: 24,190

REFERENCE/DOCKET NUMBER: A-58783

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-781-1989

TELEFAX: 415-398-3249

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 60 amino acids

TYPE: amino acid

STRANDEDNESS: single

MOLECULE TYPE: linear

FEATURE:

NAME/KEY: Protein

LOCATION: 1-60

OTHER INFORMATION: /note- "Xaa refers to stop codon in

OTHER INFORMATION: the open reading frame."

SEQUENCE 60 AA: 7131 MW: 22960 CN:

Query Match 2.2%: Score 148; DB 1; Length 60;  
Best Local Similarity 47.1%; Pred. No. 3.79e+01;  
Matches 33; Conservative 15; Mismatches 22; Indels 0; Gaps 0;

Db Y V I I I D L I H F L H A N Y L N V I N Y  
Dt : : : : : : : : : : : : : : : : : :  
OY 514 TACCAATTAAGCAATATGATTTTCATTCATCTGATTTTCAAGATTAATGAT 573  
Qt Y A I M I N I D F D H P D Y F K D I N D

Db

Db

D F D



Dt 118 GATTTCGATG 127  
118 GATTTCGATG 127  
Gt 574 GATTTCGATG 583  
V F D

RESULT 12  
ID US-07-991-867B-3 STANDARD: PRT: 225 AA.

AC xxxxxx

Sequence 3, Application US/07991867B

Sequence 3, Application US/07991867B  
Patent No. 5476781

GENERAL INFORMATION:

APPLICANT: Moyer, Richard W.

APPLICANT: Hall, Richard L.

APPLICANT: Gruidl, Michael E.

TITLE OF INVENTION: No. 5476781 Entomopoxvirus Expression System

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: FL

COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/991,867B

FILING DATE: 12-DEC-1992

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO 92/14818

FILING DATE: 12-FEB-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/827,685

FILING DATE: 30-JAN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/657,584

FILING DATE: 19-FEB-1991

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.

REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: UF114.C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 904-375-8100

TELEFAX: 904-375-5800

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 225 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 225 AA; 26008 MW; 256063 CN;

Query Match 2.2%; Score 147; DB 1; Length 225;  
Best Local Similarity 38.5%; Pred. No. 4.23e+01;

Matches 42; Conservative 25; Mismatches 42; Indels 0; Gaps 0;

Dt S I N N I L N N F A I I T N V R L I D K  
555 HMSNTHAAYATHTYNAAYATTCGNATHACNNAAYGTMNGYTNATGAYAA 614  
Ct ATGCATCAAAACATCATTAATCTTGAATATCAAGATGATCGAATCAATATTGG 527  
C I K N I I N I F E I I R M I E I N I C

Dt Y N S I I S F L N I N V G T L F  
615 RTAAYAAWSNTHATHTMSNTTYTNAAYATHAAYGTMNGNACNTNTTY 663  
Ct TCATATATGCGTAATCAAGTTTATACCTTAAGAAGTGCCTATATTC 478  
H N C V I I R F I T \* K V T S I F

RESULT 13  
ID US-08-107-755A-3 STANDARD: PRT: 225 AA.

AC xxxxxx

Sequence 3, Application US/08107755A

Sequence 3, Application US/08107755A  
Patent No. 5721352

GENERAL INFORMATION:

APPLICANT: Moyer, Richard W.

APPLICANT: Hall, Richard L.

APPLICANT: Gruidl, Michael E.

TITLE OF INVENTION: No. 5721352 Entomopoxvirus Expression System

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: Florida

COUNTRY: U.S.A.

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/107,755A

FILING DATE: 19-AUG-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/827,658

FILING DATE: 30-JAN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/657,584

FILING DATE: 19-FEB-1991

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.

REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: UF114.C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (904) 375-8100

TELEFAX: (904) 372-5800

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 225 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 225 AA; 26008 MW; 256063 CN;

Query Match 2.2%; Score 147; DB 1; Length 225;  
Best Local Similarity 38.5%; Pred. No. 4.23e+01;

Matches 42; Conservative 25; Mismatches 42; Indels 0; Gaps 0;

Dt S I N N I L N N F A I I T N V R L I D K  
555 HMSNTHAAYATHTYNAAYATTCGNATHACNNAAYGTMNGYTNATGAYAA 614  
Ct ATGCATCAAAACATCATTAATCTTGAATATCAAGATGATCGAATCAATATTGG 527  
C I K N I I N I F E I I R M I E I N I C

Dt Y N S I I S F L N I N V G T L F  
615 RTAAYAAWSNTHATHTMSNTTYTNAAYATHAAYGTMNGNACNTNTTY 663





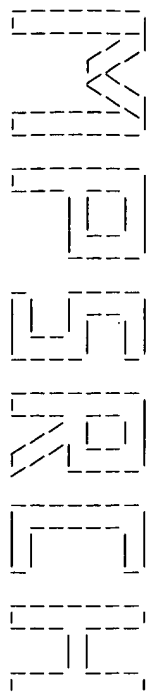






**THIS PAGE BLANK (USPTO)**





(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_an n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 10:23:08 1999; Maspar time 2454.85 Seconds  
1525.488 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-1  
Description: (1-1351) from US09103287.seq  
Perfect Score: 1351  
N.A. Sequence: 1 ANGAGTACAGGAGTTTATAT.....TTATATGTTTATATATGAG 1351  
Comp: TACTCATTCCTCAAAATATATA.....AATTATACAAATATATCTC

Scoring table: TABLE default  
Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 646147 segs, 1385953633 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb158  
1:em\_ba1 2:em\_ba2 3:em\_fun 4:em\_htg 5:em\_hum1 6:em\_hum2  
7:em\_in 8:em\_ov 9:em\_ov 10:em\_ov 11:em\_pat 12:em\_ph  
13:em\_pl 14:em\_ro 15:em\_sts 16:em\_v1

Database: genbank111

17:gb\_ba1 18:gb\_ba2 19:gb\_htg1 20:gb\_htg2 21:gb\_in1  
22:gb\_in2 23:gb\_ov 24:gb\_ov 25:gb\_pat 26:gb\_ph 27:gb\_pl1  
28:gb\_pl2 29:gb\_pl 30:gb\_pl2 31:gb\_pl3 32:gb\_ro  
33:gb\_st 34:gb\_sts 35:gb\_sy 36:gb\_un 37:gb\_v1

Statistics: Mean 11.269; Variance 6.551; scale 1.720

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description	Pred. No.
1	1282	94.9	1314 17	AF034076	Staphylococcus aureus	0.00e+00
2	308	22.8	215640 17	BSUB0016	Bacillus subtilis comp	6.77e-191
3	308	22.8	220060 18	AF008220	Bacillus subtilis rnb	6.77e-191
4	259	19.2	2320 17	BACUNAM	Bacillus subtilis UDP-	1.13e-135
5	59	4.4	7218 25	166494	Sequence 14 from paten	5.17e-18
6	41	3.0	10772 21	AF012089	Drosophila melanogaste	1.98e-07
7	38	2.8	10772 21	AF012089	Borrelia burgdorferi (	3.77e-06
8	33	2.4	15079 18	AE001180	Musella vison GT dinuc	3.84e-02
9	31	2.3	1056 23	MYU87256	Buchnera aphidicola UD	3.84e-02
10	31	2.3	6757 18	AF012886	Aquifex aeolicus secti	3.84e-02
11	31	2.3	15862 18	AE000736	Haemophilus influenzae	1.19e-01
12	29	2.2	12085 18	U32794	Oxytricha fallax 57kd	3.63e-01
13	29	2.1	354 21	OFU89259		

14	2.1	1056 23	MYU87256	Musella vison GT dinuc	3.63e-01
15	2.1	1430 28	AF015636	Epicaris impressa matur	1.09e+00
16	2.1	2100 17	LBAPREL	Lactobacillus sp. gene	1.09e+00
17	2.1	74371 31	AC005369	Homo sapiens chromosom	1.09e+00
18	2.1	216021 31	HUAC004787	Homo sapiens Chromosom	3.63e-01
19	2.1	312430 17	RPX02	Rickettsia prowazekii	1.09e+00
20	2.0	215 25	128278	Sequence 5 from patent	3.17e+00
21	2.0	1200 25	173245	Sequence 2 from patent	3.17e+00
22	2.0	1200 25	123951	Sequence 2 from patent	3.17e+00
23	2.0	1200 25	A29523	fungai stress protein.	3.17e+00
24	2.0	3885 21	PFSC03104	Plasmodium falciparum	3.17e+00
25	2.0	16157 18	AE001670	Chlamydia pneumoniae s	3.17e+00
26	2.0	112217 21	PFMAL3P8	Plasmodium falciparum	3.17e+00
27	2.0	116156 28	ATAC002343	Arabidopsis thaliana B	3.17e+00
28	1.9	215 25	128278	Sequence 5 from patent	9.05e+00
29	1.9	2921 17	MC385	Mycapricolium DNA for C	9.05e+00
30	1.9	922 17	PFASNS514	Plasmodium falciparum	9.05e+00
31	1.9	1370 31	AF052161	Homo sapiens clone 244	2.52e+01
32	1.9	1384 28	AF015643	Rupicola sprengelii	2.52e+01
33	1.9	1677 21	AB006354	Plasmodium falciparum	2.52e+01
34	1.9	2223 28	CMZCPNHF	Campylobacter ramosa chlor	9.05e+00
35	1.9	5624 28	AF043301	Candida albicans retro	2.52e+01
36	1.9	5990 17	CBDEVNA1	Carnobacterium diverge	9.05e+00
37	1.9	11680 18	HPAE00609	Helicobacter pylori 26	9.05e+00
38	1.9	32346 21	CEB0240	Caenorhabditis elegans	9.05e+00
39	1.9	36626 21	CEIC15D9	Caenorhabditis elegans	9.05e+00
40	1.9	139444 29	HS380C13	Human DNA sequence fto	9.05e+00
41	1.9	146111 19	HS571B2	Human DNA sequence ***	9.05e+00
42	1.9	163493 19	HS404H4	Human DNA sequence ***	9.05e+00
43	1.9	170417 20	AC005139	Plasmodium falciparum	9.05e+00
44	1.9	216021 31	HUAC004787	Homo sapiens Chromosom	9.05e+00
45	1.9	236120 37	AF063866	Melanoplus sanguinipes	9.05e+00

## ALIGNMENTS

LOCUS 1 AF034076 1314 bp DNA BCT 26-NOV-1997  
DEFINITION Staphylococcus aureus UDP-N-acetylmuramoyl-L-alanine synthetase  
(murC) gene, complete cds.  
ACCESSION AF034076  
VERSION 92642658  
NID AF034076.1 GI:2642658  
KEYWORDS

SOURCE  
ORGANISM Staphylococcus aureus.  
Staphylococcus aureus.  
Eubacteria; Firmicutes; low G+C gram-positive bacteria;  
Bacillaceae; Staphylococcus.

REFERENCE  
1 (bases 1 to 1314)

TITLE  
JOURNAL Cloning and sequencing of Staphylococcus aureus murC, a gene  
involved in cell wall biosynthesis  
Unpublished  
2 (bases 1 to 1314)

REFERENCE  
Lowe,A.M. and Deresiewicz,R.L.  
Submitted (11-NOV-1997) Channing Laboratory, Brigham and Women's  
Hospital and Harvard Medical School, 181 Longwood Ave, Boston, MA  
02115, USA

## FEATURES

source 1. 1314  
Location/Qualifiers  
/organism="Staphylococcus aureus"  
/db\_xref="taxon:1280"  
1. 1314  
/gene="murC"  
1. 1314  
/gene="murC"  
/function="cell wall biosynthesis"  
/note="MurC: UDP-N-acetylmuramate-alanine ligase"  
/codon\_start=1  
/transl\_table=11  
/product="UDP-N-acetylmuramoyl-L-alanine synthetase"  
/protein\_id="AAB87090.1"







[illegible]



/db\_xref="GI:2635418"  
 /db\_xref="SPRMBL:034900"  
 /translation="MIEKINIKHOFHIVLAKINLTVRKEGVIIIGSGSKTEFL  
 RCLNLEPRPEGIIISHDKYINCRPSKRVHVLKQAVFOQLHFAKTYENVM  
 EGIJIAKMKODAVAVENELRKGVLQDNLNPOSLSGOKORGIARALAHIDV  
 LLEDEPAAIDPELVGEVLEVMLEIKYTKQTMIVVHEMFARVSDQVFMDEGIV  
 EOGTPEEVRHNIKDRKQFRLKRVSPETYLEPEKHEKEPIV"  
 gene  
 complement(6425..7132)  
 /gene="ytmM"  
 /complement(6425..7132)  
 CDS  
 /function="unknown"  
 /note="similar to amino acid ABC transporter (permease)"  
 /codon\_start=1  
 /transl\_table=11

Note: remainder of annotations omitted.  
 Query Match 22.8%, Score 308, DB 17, Length 215640.  
 Best Local Similarity 62.1%; Pred. No. 6.77e-191;  
 Matches 789; Conservative 0; Mismatches 481; Indels 0; Gaps 0;

Db 49522 ATTTTGATATCCGCTCCGATGAGATGAGAGGCTTTATCATGAGCTTTTAAA 49581  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 1288 ATTTTGAATATCCTGACACCCATTAATTAACACACGATTAATGTTCTATA 1229  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 49582 CAGATGTCATCTTCTTCATCAGCTTGATTAATTTTCCCTGAGATCAGCA 49641  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 1228 CATTAATAGATCTTCAATTAATGAGCATGACCTCCATTTTATCAATTAATCTGTA 1169  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 49642 TCGACACTTCCGGCATCTCAGCGGCTGAGCGGAAATATGCAATATACACAGT 49701  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 1168 TCGTAATGCGCAGATTTTCTTAATGAGCCAAAATTTCAATAGATACAGAT 1109  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 49702 CGGCGCGCTCAGGCTTCTGCAATTCGTCAGAGACTGCTGCGGTAATATAT 49761  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 1108 CTGCTTACATTAACCTTTCGCAATTCATTTAAAAATGCTTGTCTTAGAGAAATGCT 1049  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 49762 GAGGCTGGAATACCGGCAATTTCCGATCAGATCTTGTCTTCCGCTCGATG 49821  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 1048 GTGCTGAATACTGCAACAACCTTTATGATGATTTCTTTCGCTGCTGATTTG 989  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 49822 TCACCTTATTTCTGTGGATGATGGGATATGATCATACGACTTTGCCAAGCT 49881  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 988 TAGCAGTATTTCTTGTGATGATGATCATATCATATGATTTGCAATTTG 929  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 49882 GCTTCATGATGATCGGATTTGAGCGCCCAAGGATTTGAGAGCATGCTATATGTC 49941  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 928 TAGTTTATGAAACGACGTTTAACACCAACGTTTCTTAATGCTTTTAAATTTG 869  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 49942 TGAATCAATTTTCAATATGCAATACGATACGACGCAATGATTAATCATTTG 50001  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 868 TAACATGATCTCTTAATTAATTAATGCAATTAACGATTAATTAATTAATTAAT 809  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50002 GGTGCGCGTACGCGAATTAACACTATCATAGAACGATATGCGACAAACATCAA 50061  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 808 GGTACCATATTTGAGACAGAAATGATCAATAACTACCATACATACATCAA 749  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50062 AAGTGTCCCTGCGCTTTTACGATGTTTCTGCTTAATAATCAATTTCTCCCGC 50121  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 748 AAGGATGATCTTAATCCGTAATTTGATTAATTAATTAATTAATTAATTAAT 689  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50122 TGCCATATACACACCGGAGCTTCATGATTTTCCGAGATGCTTCGCTCCG 50181  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 688 ATCCATATGATTAATTAATGAAACATCTCTTAATTAATTAATTAATTAATTAAT 629  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50182 AGGCAATATGCTTTGATGATTAAGACCATCTCTGAAAGCGTCAATATCATCTG 50241  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 628 AAGCAATATATCTTTTAAATTAATGATGATTTCTGAAATGATCAAAAACATCAT 569  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50242 CAATGCTTGAAGATATCAGGATGATCAATATTCATATGATATGATATCTG 50301  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 568 TAATATCTTGAATATATCAGGATGATCAATATTTTGTATTAATGCGTAATCAG 509  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 50302 GCTGATAGCTTAAGAAATGCGCGGATTTGCGACGCTTCAACAAAGATATCGCTG 50361  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 508 GTTATTAATTAAGAAATGAGCGCTATATACATGCGCTCAAAAAGGAAATATCACTTT 449  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50362 TTTCATTTCCCTGCGCTGTTCCGCTTCCGATTAAGAAAGACGCTGTTGGCTTTGGA 50421  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 448 CAGGCAATCCCAATACCTGTGCGCATCAATTAATTAATTAATTAATTAATTAATTAAT 389  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50422 TCACGATGACGACAGACCGGATGAGCTTTGCGGTGCGCGGTAACGCAAC 50481  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 388 TAACATGATGATTAATTAACCTGTTGTAGAGTTTACCATATGTCACAGTATACACTAG 329  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50482 TCGTGAATTTTTCATGATGATGCGCTTAATAAATTAATGATATACGATACCGGATTCCT 50541  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 328 AAGTATATGATCAATTAATCTGCTTAATAAATCAATTAATTAATTAATTAATTAATTAAT 269  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50542 CAGACATCGCTTTTCTATTTCAAGATGCGTCTGGAATGCGTTCCGCGCATCTG 50601  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 268 ATGATGATGACCTACTATTTCTTCATGCTACTGCGAATGATGATGATGATGATGATGAT 209  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50602 TCATGCGGCTTAATTAATTTCCGCTTAAGAGAAATCGTAATTTCTTTTCAA 50661  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 208 CCAATCTTCTTTATGATTAATTAACACCAATGTAATTTTATCCCTTATTTCTTA 149  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50662 GCGCTGTGCGTAATAATTTTTCATATCCGATCCTTGACAGTATATCATTAAT 50721  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 148 GAGCAACTGCTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 89  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50722 CATGAAGTATTTGGGACGAGCTATACGCTGCTTTTATCCAAACAAATGATTA 50781  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 88 CATGATGATTTGCTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 29  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50782 CAGTCAATAT 50791  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 28 GTTCATTAAT 19  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3  
 LOCUS AF008220 22060 bp DNA BCT 04-FEB-1998  
 DEFINITION Bacillus subtilis rmb-dnab genomic region.  
 ACCESSION AF008220  
 NID 92293135  
 VERSION AF008220.1 GI:2293135  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Bacillus subtilis.  
 Bacillus subtilis.  
 Eubacteria; Firmicutes; Low G+C gram-positive bacteria;  
 Bacillaceae; Bacillus.  
 1 (bases 1 to 7430)  
 Green, C.J., Stewart, G.C., Hollis, M.A., Vold, B.S. and Botl, K.F.  
 Nucleotide sequence of the Bacillus subtilis ribosomal RNA operon,  
 rmb.  
 JOURNAL  
 Gene 37 (1-3), 261-266 (1985)  
 MEDLINE  
 86031361  
 REFERENCE  
 2 (bases 153210 to 153762)  
 Connors, M.J., Mason, J.M. and Setlow, P.  
 Cloning and nucleotide sequencing of genes for three small,  
 acid-soluble proteins from Bacillus subtilis spores  
 J. Bacteriol. 166 (2), 417-425 (1986)  
 JOURNAL  
 J. Bacteriol. 166 (2), 417-425 (1986)  
 MEDLINE  
 86195826  
 REFERENCE  
 3 (bases 213161 to 218473)  
 Ogasawara, N., Moriya, S., Maeda, P.G. and Yoshikawa, H.  
 Nucleotide sequence and organization of dnaB gene and neighbouring  
 genes on the Bacillus subtilis chromosome  
 Nucleic Acids Res. 14 (24), 9989-9999 (1986)  
 JOURNAL  
 Nucleic Acids Res. 14 (24), 9989-9999 (1986)  
 MEDLINE  
 87117549  
 REFERENCE  
 4 (bases 200404 to 201481)  
 Seki, T., Yoshikawa, H., Takahashi, H. and Saito, H.  
 Cloning and nucleotide sequence of phoP, the regulatory gene for  
 alkaline phosphatase and phosphodiesterase in Bacillus subtilis  
 J. Bacteriol. 169 (??), 2913-2916 (1987)  
 JOURNAL  
 J. Bacteriol. 169 (??), 2913-2916 (1987)  
 MEDLINE  
 87250247



REFERENCE 5 (bases 201234 to 203212)  
 AUTHORS Seki, T., Yoshikawa, H., Takahashi, H. and Saito, H.  
 TITLE Nucleotide sequence of the *Bacillus subtilis* *phoR* gene  
 JOURNAL J. Bacteriol. 170 (12), 5935-5938 (1988)  
 MEDLINE 89053932  
 REFERENCE 6 (bases 142232 to 144147)  
 AUTHORS Grundy, F.J. and Henkin, T.M.  
 TITLE Cloning and analysis of the *Bacillus subtilis* *rpsD* gene, encoding ribosomal protein S4  
 JOURNAL J. Bacteriol. 172 (11), 6372-6379 (1990)  
 MEDLINE 91035248  
 REFERENCE 7 (bases 133624 to 134990)  
 AUTHORS Henkin, T.M., Grundy, F.J., Nicholson, N.L. and Chambliss, G.H.  
 TITLE Carboxylate repression of alpha-amylase gene expression in *Bacillus*  
*subtilis* involves a trans-acting gene product homologous to the  
*Escherichia coli* *lacI* and *galP* repressors  
 JOURNAL Mol. Microbiol. 5 (3), 575-584 (1991)  
 MEDLINE 91260441  
 REFERENCE 8 (bases 140810 to 142610)  
 AUTHORS Henkin, T.M., Glass, B.L. and Grundy, F.J.  
 TITLE Analysis of the *Bacillus subtilis* *tyrS* gene: conservation of a regulatory sequence in multiple tRNA synthetase genes  
 JOURNAL J. Bacteriol. 174 (4), 1299-1306 (1992)  
 MEDLINE 92138624  
 REFERENCE 9 (bases 217570 to 220060)  
 AUTHORS Putzer, H., Gendron, N. and Grunberg-Manago, M.  
 TITLE Co-ordinate expression of the two threonyl-tRNA synthetase genes in *Bacillus subtilis*: control by transcriptional antitermination involving a conserved regulatory sequence  
 JOURNAL EMBO J. 11 (8), 3117-3127 (1992)  
 MEDLINE 92347349  
 REFERENCE 10 (bases 134990 to 141290)  
 AUTHORS Grundy, F.J., Waters, D.A., Takaya, T.Y. and Henkin, T.M.  
 TITLE Identification of genes involved in utilization of acetate and acetoin in *Bacillus subtilis*  
 JOURNAL Mol. Microbiol. 10 (2), 259-271 (1993)  
 MEDLINE 95020526  
 REFERENCE 11 (bases 162129 to 164080)  
 AUTHORS Grundy, F.J., Waters, D.A., Allen, S.H. and Henkin, T.M.  
 TITLE Regulation of the *Bacillus subtilis* acetate kinase gene by *CcpA*  
 JOURNAL J. Bacteriol. 175 (22), 7348-7355 (1993)  
 MEDLINE 94042910  
 REFERENCE 12 (bases 7009 to 15526)  
 AUTHORS Kiel, J.A., Boels, J.M., Beldman, G. and Venema, G.  
 TITLE Glycogen in *Bacillus subtilis*: molecular characterization of an operon encoding enzymes involved in glycogen biosynthesis and degradation  
 JOURNAL Mol. Microbiol. 11 (1), 203-218 (1994)  
 MEDLINE 94195107  
 REFERENCE 13 (bases 196487 to 200620)  
 AUTHORS Jin, S. and Sonenshein, A.L.  
 TITLE Identification of two distinct *Bacillus subtilis* citrate synthase genes  
 JOURNAL J. Bacteriol. 176 (15), 4669-4679 (1994)  
 MEDLINE 94321340  
 REFERENCE 14 (bases 16985 to 19588)  
 AUTHORS Abe, A., Koide, H., Kohno, T. and Watabe, K.  
 TITLE A *Bacillus subtilis* spore coat polypeptide gene, *cots*  
 JOURNAL Microbiol. 141 (Pt 6), 1433-1442 (1995)  
 MEDLINE 95400496  
 REFERENCE 15 (bases 131934 to 133970)  
 AUTHORS Bolotin, A., Khazak, V., Stoyanova, N., Ratmanova, K., Yomantas, Y. and Kozlov, Y.  
 TITLE Identical amino acid sequence of the *aroA(G)* gene products of *Bacillus subtilis* 168 and *B. subtilis* Marburg strain  
 JOURNAL Microbiology 141 (Pt 9), 2219-2222 (1995)  
 MEDLINE 96118703  
 REFERENCE 16 (bases 25258 to 31212)  
 AUTHORS Rowland, B., Hill, K., Miller, P., Driscoll, J. and Taber, H.  
 TITLE Structural organization of a *Bacillus subtilis* operon encoding menaquinone biosynthetic enzymes  
 JOURNAL Gene 167 (1-2), 105-109 (1995)  
 MEDLINE 96144257

REFERENCE 17 (bases 196487 to 200620)  
 AUTHORS Jin, S., De Jesus-Berrios, M. and Sonenshein, A.L.  
 TITLE *Bacillus subtilis* malate dehydrogenase gene  
 JOURNAL J. Bacteriol. 178 (2), 560-563 (1996)  
 MEDLINE 96134995  
 REFERENCE 18 (bases 129888 to 132207)  
 AUTHORS Varon, D., Brody, M.S. and Price, C.W.  
 TITLE *Bacillus subtilis* operon under the dual control of the general stress transcription factor sigma B and the sporulation transcription factor sigma H  
 JOURNAL Mol. Microbiol. 20 (2), 339-350 (1996)  
 MEDLINE 96310371  
 REFERENCE 19 (bases 81540 to 91690)  
 AUTHORS Bower, S., Perkins, J.B., Vocum, R.R., Howitt, C.L., Rahaim, P. and Pero, J.  
 TITLE Cloning, sequencing, and characterization of the *Bacillus subtilis* biotin biosynthetic operon  
 JOURNAL J. Bacteriol. 178 (14), 4122-4130 (1996)  
 MEDLINE 96312354  
 REFERENCE 20 (bases 49093 to 51682)  
 AUTHORS Vocum, R.R., Perkins, J.B., Howitt, C.L. and Pero, J.  
 TITLE Cloning and characterization of the *metE* gene encoding S-adenosylmethionine synthetase from *Bacillus subtilis*  
 JOURNAL J. Bacteriol. 178 (15), 4604-4610 (1996)  
 MEDLINE 96345628  
 REFERENCE 21 (bases 100760 to 102298)  
 AUTHORS Kappes, R.M., Kempf, B. and Bremer, E.  
 TITLE Three transport systems for the osmoprotectant glycine betaine operate in *Bacillus subtilis*: characterization of *OpdJ*  
 JOURNAL J. Bacteriol. 178 (17), 5071-5079 (1996)  
 MEDLINE 96359357  
 REFERENCE 22 (bases 1 to 220060)  
 AUTHORS Lapidus, A., Galleron, N., Sorokin, A. and Ehrlich, S.D.  
 TITLE Sequencing and functional annotation of the *Bacillus subtilis* genes in the 200 kb *rmb-dnaB* region  
 JOURNAL Microbiology 143 (Pt 11), 3431-3441 (1997)  
 MEDLINE 98048467  
 REFERENCE 23 (bases 1 to 220060)  
 AUTHORS Lapidus, A., Galleron, N., Sorokin, A. and Ehrlich, S.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-JUN-1997) Laboratoire de Genetique Microbienne, INRA, Domaine de Vilvert, Jouy-en-Josas cedex 78352, France  
 FEATURES  
 source  
 1..220060  
 /organism="Bacillus subtilis"  
 /db\_xref="taxon:1423"  
 485..2034  
 /product="16S ribosomal RNA"  
 2203..5129  
 /product="23S ribosomal RNA"  
 5185..5300  
 /product="5S ribosomal RNA"  
 5322..5397  
 /note="codon recognized: GTA"  
 5430..5504  
 /product="tRNA-Val"  
 5543..5618  
 /product="tRNA-Thr"  
 /note="codon recognized: ACA"  
 5629..5715  
 /product="tRNA-Lys"  
 5721..5795  
 /product="tRNA-Leu"  
 /note="codon recognized: GGC"  
 /product="tRNA-Gly"  
 5810..5895  
 /note="codon recognized: TTA"  
 /product="tRNA-Leu"  
 5905..5981  
 /note="codon recognized: CGT"  
 /product="tRNA-Arg"







[illegible]

REFERENCE	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE
133	CGAGCTGTTCTTAAATGCGGAGGAGGAAAGGCGGAGGAAATATAAAACGCGCAATCTTG	192						
404	AGACTTATTTTAAATTTGATGATGCGACAGGATATGGATTTGCTTGAAGACTGATTTTTCG	463						
193	TGTTTGAAGGCGGATATACCGCGGCACTTTCTTAAGTATTCAGCCAGATTATTCGATT	252						
464	CTTTTGAAGCATGATATATAGACGTCCTTTTAAAGTTATTAACCTGATTACGCAATTA	523						
253	TGACGAATATGATTTGATTCACCTTGATTTACTTCTCAAGCATTCAGATGATTTAGC	312						
524	TGACAAATATTTGATTTGATTCATCTGATTTATTTCAAAATATTAATGATTTTGTATG	583						
313	CTTTCAAGAGATGGGCTTCAGTCACAAAGGCAATTTTCCCTGCGGAGCGATAGC	372						
584	CATTTCAGAAATATGGCACATATATGTTTAAAAAAGTATTTATCTTGGGGTGATGATGAAC	643						
373	ATCTGCCGAATATCATAGCGACAGCTCCGCGTGTGTATTATGCGACGGGGAGAAATATG	432						
644	ATCTAGCTAAATATGAAGCAGATGTTCCAAATTTATTTACATGATTTTAAAGTTGGGATG	703						
433	ATTTTCAAGCAGAAACATCTGTAAAAAGCAGGAGGACAACTTTTGATGTCCTTTC	492						
704	ACATTATGCTCAAAATATTCAAATATACGATTAAGGATGACTCTTTTGATGATGATGG	763						
493	GCAATAGCTTCATGATACGTTTATATTTCTGCGGACGCGCACACCATGATTTAACT	552						
764	ATGTGTAGTTTATGATCACTCTCTCTTCACAAATATGTCACATATACGATTTAAATG	823						
553	CATTGGGCGGATCATCGCTTATGCGCATTTATGAAGAAATTTATTCAGCATCATTAAGCATG	612						
824	CATTAGCTGTAATTTGCGATTTAGTTATTTTGAAGAGCTAGATGTTACAAATATTTAAGAG	883						
613	CTCTCAATCTCTTGGGGCGTCAAAAGCAGATTTCAATGAGAAGCACTTGGGATCAAG	672						
884	CATTAGAAACGTTGGTGGTGTTAAAGCTGTTTCAATGAATCTCAATTTGCAATCAAG	943						
673	TGCGATGATGATGACTATGCCCATCTCCGACAGAAATATAAAGTGACATTCAGGCGGCA	732						
944	TTATTTGATGATGATTAAGCACACCATTCAGAAAGAAATTTGCTCCAAATTTGACACAGAC	1003						
733	GACGAGATTAACCGATCGGGAATTTGCGGGTATTCAGGCTCATACATTACCCGGA	792						
1004	GAAAGAAATATCCCATTAAGAAAGTGTGTCGATTTATTAACCACACACTTCTCTAGAA	1066						
793	CGCAGACATTCCTTGACGAATTTGCAGAAAAGCCTGAAGCGGCGCCAGACTGTGTATTAT	852						
1064	CACAGACATTTTAAATGATTTTGCAGAAAAGTTATGTAAGACAGATGCTGATTTCTTAT	1123						
853	GCGATATTTTGGCTCAGCCCGTGAGAAATGCCGGAAGCTGACGATCGGTATCTCCAG	912						
1124	GTTGAAATTTTGGCTCAATTAGAGAAAATTTTGGCCATTACGATACAAAGTTTATATG	1183						
913	GAAATATTCATTAATGCCAACCTGATTGAAGAAAGATGACACATCTGTTTAAAGGCTATG	972						
1184	ATTAATATTTGAGGTGCATCGTTATTAATGAATATCTTTATTAATGATTAAGAACAAATTTG	1243						
973	ATAAAGCGGTTCATCTTCATGATGGAGAGAGATATCCAAAAAT	1017						
1244	ATAATGCTGTTGTTTATTTATATGGGAGGAGGATATTCAAAAAT	1288						
RESULT	5							
LOCUS	166494	7218 bp	DNA					
DEFINITION	Sequence 14 from patent US 5670367.							
ACCESSION	166494							
NID	92724471							
KEYWORDS	166494.1	GI:2724471						
SOURCE	Unknown.							
ORGANISM	Unknown.							
REFERENCE	1 (bases 1 to 7218)							
AUTHORS	Dorner, F., Scheffligner, F. and Falkner, F. Gunter.							



